

A DNA sequence encoding human β -casein (a human milk protein) and analogues of said DNA sequence which encodes a polypeptide having the calcium binding activity of human β -casein or opioid activity or angiotension converting enzyme (ACE) inhibitory activity. The polypeptide encoded by said DNA sequence or analogues may be used in the production of recombinant human β -casein either by means of a prokaryotic or an eukaryotic production system, or more advantageously by means of production in transgenic non-human mammals such as cows. The recombinant human β -casein is a useful constituent of infant formulae used for feeding infants as a substitute for human milk, i.e. maternal milk. The recombinant human β -casein makes it possible to prepare an infant formula having a closer similarity to human milk and thus an improved nutritional and biological value.

HUMAN β -CASEIN, PROCESS FOR PRODUCING IT AND USE THEREOF.

The present invention relates to a DNA sequence encoding the human milk protein β -casein. More specifically, the DNA sequence encodes a polypeptide having the amino acid sequence shown in Figure 1 below.

5 The DNA sequence is advantageously used in the production of recombinant human β -casein, either by means of a prokaryotic or an eukaryotic production system, or more advantageously by means of production in transgenic non-human mammals such as cows. The main use of the recombinant human β -casein is as a constituent of infant formulae
10 used for feeding infants as a substitute for human milk. When used as a constituent of infant formulae, it is contemplated that the recombinant human β -casein provides a substantial improvement of the nutritional and biological value of the formulae in that a closer similarity to human milk is obtained.

15

BACKGROUND OF THE INVENTION

It is well known that human milk-feeding is considered superior to formula-feeding for infants. Not only does human milk provide a well-balanced supply of nutrients, but it is also easily digested by the
20 infant. Thus, several biologically active components which are known to have physiological functions in the infant are either a constituent of human milk or produced during the digestion thereof, including components involved in the defense against infection and components facilitating the uptake of nutrients from human milk.

25 In spite of the great efforts which have been invested in preparing infant formulae, it has not been possible to produce a formula which to any substantial extent has the advantageous properties of human milk. Thus, infant formula, often prepared on the basis of cow milk, is generally incompletely digested by the infant and is lacking sub-
30 stances known to have effect on the physiological functions of the infant. In order to obtain an infant formula with a nutritional value similar to human milk, a number of additives including protein fragments, vitamins, minerals etc., which are normally formed or taken up during the infant's digestion of human milk, are included in the for-
35 mula with the consequent risk of posing an increased strain on and

possible long-term damage of important organs such as liver and kidney. Another disadvantage associated with the use of cow milk-based formulae is the increased risk for inducing allergy in the infant against bovine proteins.

- 5 As an alternative to cow milk-based infant formulae, human milk obtainable from so-called milk banks has been used. However, feeding newborn infants with human milk from milk banks has in the recent years to an increasing extent been avoided, because of the fear for the presence of infective agents such as HIV and CMV in human milk.
- 10 In order to destroy the infective agents in human milk it has become necessary to pasteurize the milk before use. However, by pasteurization the nutritional value and the biological effects of the milk components are decreased and human milk is used to a still lesser extent.
- 15 Human milk differs markedly from that of other mammalian species, including cows, in that it contains a lower over-all protein content and lower ratio of casein/whey as well as a different protein composition. For instance, the casein subclasses of human milk comprises only β -casein and κ -casein, whereas the bovine casein subclasses are
- 20 α -casein, β -casein, and κ -casein (Miller et al.). Also the amino acid sequences of human milk protein differ from that of other mammalian milk proteins.

- β -Casein is a phosphorylated protein which is present in milk of several species including humans in which it is the major casein
- 25 subunit. This protein - or its digested fragments - is believed to enhance calcium absorption by chelating calcium to its phosphorylated residues and thereby keeping it in an absorbable form. Human β -casein is easily digested by newborn infants and the digestive products have been found to play an important role in the calcium uptake, and thus
- 30 in the mineralization of the skeleton. A digestion product (β -casomorphin) of human β -casein has been found to have opioid activity and may be involved in the sleeping patterns of breast-fed infants.

It would be desirable to be able to prepare an infant formula with a composition closer to that of human milk and thus avoid the above

disadvantages associated with bovine milk-based infant formula, e.g. a formula comprising human milk proteins. However, this would require that human milk proteins are obtainable in large quantities. Although human milk proteins may be purified directly from human milk, this is not a realistic and sufficiently economical way to obtain the large quantities needed for large scale formula production, and other methods must be developed before an infant formula comprising human milk proteins may be prepared.

So far, little detailed characterization of human milk proteins, e.g. in terms of amino acid sequences, has been made. Greenberg et al. reported the isolation and purification of native human β -casein and the amino acid sequence thereof. The total composition of the native human β -casein was stated to be 212 amino acids and various phosphorylation sites were identified in the sequence.

Several milk protein genes, primarily from rodents or dairy animals, have been cloned and sequenced, cf. references 11-21 numbered in the List of References below, but knowledge of the genes encoding human milk proteins is still sparse. Recently, Menon and Ham disclosed the isolation and sequencing of a partial cDNA clone encoding human β -casein. The clone included the coding sequence corresponding to amino acid residues 146-212 of the mature protein and the entire 3' non-coding region. The deduced partial amino acid sequence was compared with the sequence of the native protein reported by Greenberg et al. and several amino acid differences were found between the two amino acid sequences. Menon and Ham have not, however, succeeded in isolating the entire cDNA encoding human β -casein, nor in providing a DNA sequence useful for the production of a polypeptide comprising a substantial part (of more than 67 amino acid residues) of the human β -casein sequence. This is accomplished for the first time by the present inventors (Lönnerdal et al, published on 20 August, 1990, which is hereby incorporated by reference).

A
d
a

Lönnerdal

Anders

BRIEF DESCRIPTION OF THE INVENTION

It is an object of the present invention to provide a means for producing recombinant human β -casein in a high yield and at a realistic price.

- 5 Accordingly, in one aspect the present invention relates to a DNA sequence encoding a polypeptide comprising the amino acid sequence shown in Figure 1 or

an analogue of said DNA sequence which

- 10 1) hybridizes with the DNA sequence shown in Figure 1 or a specific part thereof under the stringent hybridization conditions or
- 2) encodes a polypeptide, the amino acid sequence of which is at least 85% homologous with the amino acid sequence shown in Figure 1, or

- 15 3) constitutes an effective subsequence of said DNA sequence,

which encodes a polypeptide having the calcium binding activity of human β -casein, or having opioid activity, or having angiotensin converting enzyme (ACE) inhibitory activity, or a polypeptide having a combination of any of these activities.

- 20 The DNA sequence of the invention was determined on the basis of a cDNA clone isolated from a human mammary gland cDNA library using hybridization with a specific 42-mer oligonucleotide probe prepared on the basis of amino acid residues 117-130 of the human β -casein sequence as disclosed by Greenberg et al. The procedure used for
- 25 isolating the human β -casein cDNA sequence is outlined in Example 1 below.

The stringent hybridization conditions referred to above are to be understood in their conventional meaning, i.e. that hybridization is

carried out at 67°C in 2xSSC and final washing at 67°C in 1xSSC using the method specified in the "Definition" part of the Examples below.

5 The term "homologous" is used here to illustrate the degree of identity between the amino acid sequence of a given polypeptide and the amino acid sequence shown in Figure 1. The amino acid sequence to be compared with the amino acid sequence shown in Figure 1 may be deduced from a DNA sequence, e.g. obtained by hybridization as defined above, or may be obtained by conventional amino acid sequencing methods. The degree of homology is preferably determined on the amino acid sequence of a mature polypeptide, i.e. without taking any leader sequence into consideration. It is preferred that the degree of homology is at least 90%, such as 95% or even 98% with the amino acid sequence shown in Figure 1.

15 The term "effective subsequence" as used above refers to a subsequence which encodes a peptide being at least partially functional with respect to the activities of human β -casein as defined above. The subsequence may be the result of a truncation at either end of the DNA sequence or of the removal of one or more nucleotides or nucleotide sequences within DNA sequence. Preferably, when encoding a peptide having the opioid or the ACE-inhibitory activity of human β -casein, the effective subsequence comprises at least 15 nucleotides such as at least 20 nucleotides.

25 The term "calcium-binding activity" denotes the capability of the polypeptide of the invention to bind calcium and may be determined by equilibrium dialysis or a similar technique. The term "opioid activity" denotes the peptide's opiate-like effects and capability of the peptide to bind to opiate receptors (opiate receptor affinity). The "opioid activity" is determined as disclosed by Brantl, 1984 and Migliori-Samour et al. The term "ACE-inhibitory activity" denotes the capability of the peptide to inhibit the angiotensin converting enzyme (ACE) and has important indications for the treatment of heart disorders. The ACE-inhibitory effect is determined by use of a method as disclosed by Maruyama et al. and Kohmura et al.

In this connection, it should be noted that the terms "calcium-binding activity", "opioid activity" and "ACE-inhibitory activity" and related terms should be understood to be qualitative and/or quantitative that is, relating first of all to the nature of the activity, such as the nature of the calcium binding activity, and/or to the level of activity of the polypeptide as determined with reference to human β -casein. Concerning the ACE-inhibitory effect or the opioid effect, these are also of the same qualitative nature as ascribed in the literature to digestive fragments of human β -casein having ACE-inhibitory activity or opioid activity, respectively.

In this connection the term "digestive fragment" refers to the peptide fragment(s) which, in nature, are generated during the digestion of human β -casein by the infant fed on human milk. Such fragments may be prepared, e.g. by cleavage of recombinant human β -casein, by expression from DNA sequences encoding such fragments, or by use of conventional peptide synthesis.

In another aspect the present invention relates to a polypeptide produced by the DNA sequence of the invention, preferably a recombinant human β -casein protein having the amino acid sequence shown in Figure 1 or a variant thereof having the calcium-binding activity of human β -casein or a subsequence of the amino acid sequence shown in Figure 1 or the variant having ACE-inhibitory activity or opioid activity of a digestive fragment of human β -casein. The variant and the subsequence is further defined below.

In yet another aspect the present invention provides a mammalian expression system comprising a DNA sequence encoding human β -casein inserted into a gene encoding a milk protein of a non-human mammal so as to form a hybrid gene which is expressible in the mammary gland of an adult female of a mammal harbouring said hybrid gene so that human β -casein is produced when the hybrid gene is expressed.

In yet a further aspect, the present invention relates to a method of producing a transgenic non-human mammal capable of expressing human β -casein, comprising injecting a mammalian expression system as defined above into a fertilized egg or a cell of an embryo of a mammal

so as to incorporate the expression system into the germline of the mammal and developing the resulting injected fertilized egg or embryo into an adult female mammal.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

5 The cDNA sequence shown in Figure 1 has an overall length of 1065 bp, including the poly(A) tail. The open reading frame begins from the first nucleotide "A" at the 5'-end and codes for a signal peptide of 15 amino acids and the mature casein consisting of 210 amino acid residues. The coding cDNA of human β -casein is flanked by one 390 bp
10 non-coding region at the 3'-terminal end and one of 48 bp at the 5'-terminus. The size of the human β -casein cDNA shown in Figure 1 is similar to that of ovine β -casein (Provot et al.).

The amino acid sequence deduced from the nucleotide sequence shown in Figure 1 shows some discrepancies compared to that reported by Green-
15 berg et al. Thus, the deduced amino acid sequence is constituted by 210 amino acid residues and no codons for the amino acids at positions 19 and 207 of the 212 amino acid sequence reported by Greenberg et al. are found. Other discrepancies are found at positions 15 (Thr
20 instead of Pro), 32 (Gly vs. Thr), 34 (Glu vs. Gln), 104 (Ser vs. Gln), 133 (Leu vs. Ser), 158 (Gln vs. Glu), 167 (Gln vs. Glu), 169 (Val vs. Leu), 173 (Gln vs. Val), 192 (Thr vs. Pro), 198 (Thr vs. Pro), 199 (Gln vs. Glu), and 201-206 (Leu-Ala-Pro-Val-His-Asn vs. Ser-Thr-Thr-Glx-Ala-Asz-His).

A relatively large degree of homology, 45-62%, is found between human
25 β -casein and the corresponding proteins from other species. Furthermore, the 15 residues signal peptide of human β -casein is identical to those of bovine, ovine and rabbit β -caseins and, except for 1 residue, identical to those of rat and mouse β -caseins. By comparison
30 between the DNA sequences encoding β -casein of different origins, single base substitutions between species, which usually occur at the wobble position and thus code for the same amino acid, have been found.

A consensus polyadenylation recognition signal AAUAAA is located 16 nucleotides upstream from the poly(A) tail. An 11 nucleotides long motif (bp 823-833; TTTATTTATTT) which might be involved in the stabilization of the mRNA, is also found and corresponds to a sequence found in connection with β -casein genes of other origins (Provot et al.).

An interesting DNA sequence translatable into a human β -casein polypeptide is a sequence comprising a human β -casein gene. Accordingly, in a further aspect, the present invention relates to a DNA sequence comprising a human β -casein gene or an effective subsequence thereof containing elements capable of expressing a polypeptide having the activity of human β -casein or a digestive fragment thereof, or an analogue of said DNA sequence which

1) hybridizes with the DNA sequence shown in Figure 1 or a specific part thereof under stringent hybridization conditions or

2) encodes a polypeptide, the amino acid sequence of which is at least 85% homologous with the amino acid sequence shown in Figure 1, or

3) constitutes an effective subsequence of said DNA sequence,

which encodes a polypeptide having the calcium binding activity of human β -casein, or opioid activity, or angiotensin converting enzyme (ACE) inhibitory activity or a combination of any two or three of these activities.

In the present context, the term "gene" is used to indicate a DNA sequence which is involved in producing a polypeptide chain and which includes regions preceding and following the coding region (5'-upstream and 3'-downstream sequences) as well as intervening sequences, the so-called introns, which are placed between individual coding segments (so-called exons) or in the 5'-upstream or 3'-downstream region. The 5'-upstream region comprises a regulatory sequence which controls the expression of the gene, typically a promoter. The 3'-

downstream region comprises sequences which are involved in termination of transcription of the gene and optionally sequences responsible for polyadenylation of the transcript and the 3' untranslated region. The term "effective subsequence" of the gene is to be understood in the same manner as defined above in connection with the DNA sequence.

The hybridization may be carried out as described in the "Definition" part of the Examples below, preferably on the basis of a probe comprising the coding part of the DNA sequence shown in Figure 1 below. The terms "homologous" and "effective subsequences" are used in a similar manner as that defined above.

Preferably, the polypeptide encoded by the analogue of the DNA sequence is at least 90% homologous, such as at least 95% or even 98% homologous with the amino acid sequence shown in Figure 1.

The β -casein gene or part thereof may be isolated and sequenced using the procedure described in Example 1C below. In Example 1B, different genetic variants of the human β -casein gene are discussed.

An example of a specific analogue of the DNA sequence of the invention is a DNA sequence which comprises the DNA sequence shown in Sequence Listing 2 and particularly adapted for expression in *E. coli*. This DNA sequence is one which, when inserted in *E. coli* together with suitable regulatory sequences, results in the expression of a polypeptide having the amino acid sequence shown in Figure 1. Thus, this DNA sequence comprises specific codons recognized by *E. coli*. The preparation of this DNA sequence is described in Example 2.

As mentioned above, the DNA sequence shown in Figure 1 encodes a polypeptide comprising the functional domain/domains of human β -casein as well as the signal peptide naturally associated therewith. While the presence of a signal peptide in most cases is a prerequisite for allowing the polypeptide expressed from the DNA sequence to be transported out of the cell in which it is produced, the nature and origin of the particular signal peptide to be used may vary and

need not be the signal peptide naturally associated with the human β -casein.

In accordance herewith, a particularly interesting DNA sequence of the invention is a DNA sequence which encodes a polypeptide comprising amino acids 1-210 shown in Figure 1, i.e. the amino acids corresponding to the mature human β -casein.

Human β -casein is highly phosphorylated at serine and threonine residues close to the N-terminal end, and this phosphorylated part of β -casein is believed to give the molecule its capacity to bind calcium and thus to participate in micelle formation. Thus, the N-terminal part of the molecule is highly important in terms of the biological activity of β -casein. In accordance herewith, an important subsequence of the DNA sequence of the invention is a DNA sequence which at least encodes the first part of the amino acid sequence shown in Figure 1, especially the part comprising amino acids 1-12. However, also other parts of the molecule may be important in connection with the calcium-binding activity of human β -casein.

Further interesting subsequences of the nucleotide and deduced amino acid sequences shown in Figure 1 are discussed in the following. A nucleotide sequence encoding a heptapeptide (VPYPQRA) expected to have ACE-inhibitory activity is found at amino acid residues 167-173 of the amino acid sequence shown in Figure 1. The expected ACE-inhibitory activity is based on the fact that a similar β -casein peptide having such activity and comprising 6 of the above listed 7 amino acid residues has been found to have ACE-inhibitory activity (Maruyama et al.). In this connection, also the pentapeptide corresponding to positions 168-172 of the amino acid sequence shown in Figure 1 is contemplated to have ACE-inhibitory activity, which pentapeptide in nature is contemplated to be a cleavage product of the heptapeptide above. Also a peptide constituted by a part of the amino acids 39-52 of the sequence shown in Figure 1, preferably a heptapeptide, a hexapeptide or a heptapeptide, is contemplated to have ACE-inhibitory activity. This expectation is based on the result of an analysis of the ACE-inhibitory activity of synthetic peptides having a similar

amino acid sequence as peptides found in the above cited region of human β -casein (Kohmura et al.).

Peptides having opioid activity and immune stimulatory activity may be determined on the basis of the disclosure of Brantl.

5 In a further aspect, the present invention relates to a modified DNA sequence which differs from a DNA sequence of the invention as defined above in that at least one nucleotide has been deleted, substituted or modified or at least one additional nucleotide has been inserted so as to result in a DNA sequence which encodes a polypeptide
10 having a calcium-binding activity which is similar to, increased or decreased as compared to the calcium-binding activity of human β -casein. Other interesting modifications result in peptides having a opioid activity or ACE-inhibitory activity.

The polypeptide encoded by the modified DNA sequence normally has an
15 amino acid sequence which is different from the amino acid sequence of the human β -casein. It will be understood that a modified DNA sequence of the invention will be of importance in the preparation of novel polypeptides having a modified activity as compared to human β -casein or digestive fragments thereof or other similarly important
20 activities.

When "substitution" is performed, one or more nucleotides in the full nucleotide sequence are replaced with one or more different nucleotides; when "addition" is performed, one or more nucleotides are added at either end of the full nucleotide sequence; when "insertion"
25 is performed one or more nucleotides within the full nucleotide sequence is inserted; and when "deletion" is performed one or more nucleotides are deleted from the full nucleotide sequence whether at either end of the sequence or at any suitable point within it.

A modified DNA sequence may be obtained by well-known methods, e.g.,
30 by use of site-directed mutagenesis.

An example of an important modified DNA sequence of the invention is a DNA sequence in which additional codons encoding serine or threo-

nine residues have been inserted so as to result in a modified DNA sequence encoding a polypeptide having an increased number of residues to be phosphorylated. The additional residues may be inserted either by being added at either end or within a DNA sequence of the invention or by replacing one or more non-serine or non-threonine codons present in a DNA sequence of the invention. A polypeptide encoded by such a modified DNA sequence is contemplated to have a higher degree of phosphorylation and thus an increased calcium-binding activity as compared to native human β -casein. The polypeptide produced from such a modified DNA sequence may be used as a nutrition supplement for populations which need to increase their calcium uptake, e.g. premature infants, women and old people.

Similarly, when a reduced calcium uptake is of interest, a modified DNA sequence of the invention encoding a polypeptide which has a lower number of residues to be phosphorylated than the polypeptide, the amino acid sequence of which is shown in Figure 1, is of interest. Here, one strategy would be to remove one or more codons for serine or threonine residues or replace one or more of such codons by non-serine or non-threonine residues.

Another example of an interesting modified DNA sequence is a DNA sequence which encodes the amino acid sequence of a naturally occurring human β -casein variant having an amino acid sequence different from the one shown in Figure 1 such as the variant disclosed by Greenberg et al. discussed above or other genetic variants as discussed in Example 1B. For this purpose, site-directed mutagenesis would be carried out using specific oligonucleotide probes conferring an exchange/removal of the relevant amino acid residues.

Another important use of a DNA sequence of the invention as defined above is in the preparation of a fusion protein comprising on the one hand a polypeptide comprising the amino acid sequence shown in Figure 1 or an analogue or subsequence thereof as defined above and on the other hand a polypeptide of another origin, e.g. a polypeptide or peptide part of another milk protein, e.g. a human milk protein such as α -lactalbumin, or a non-human milk protein such as a bovine or ovine milk protein such as bovine β -casein. The fusion protein may be

prepared by fusing a DNA sequence of the invention with a DNA sequence encoding the other part of the fusion protein and the proper regulatory sequences in a manner which allows the expression of the fusion protein to occur.

5 The DNA sequences of the invention explained herein may comprise natural as well as synthetic DNA sequences, the natural sequence typically being derived directly from cDNA or genomic DNA, normally of mammalian origin, e.g. as described below. A synthetic sequence may be prepared by conventional methods for synthetically preparing
10 DNA molecules, e.g. using the principles solid or liquid phase peptide synthesis. Of course, also the DNA sequence may be of mixed cDNA and genomic, mixed cDNA and synthetic and mixed genomic and synthetic origin.

The terms "sequence", "subsequence", "analogue" and "polypeptide" as
15 used herein with respect to sequences, subsequences, analogues and polypeptides according to the invention should of course be understood as not comprising these phenomena in their natural environment, but rather, e.g., in isolated, purified, *in vitro* or recombinant form. When reference is made to a DNA sequence of the invention this
20 should be understood to include "analogues", "subsequences" and "modified sequences" as defined above. Similarly, when reference is made to "a polypeptide of the invention" this should be understood to include any of the polypeptides defined in the following.

In another important aspect, the present invention relates to a
25 polypeptide encoded by a DNA sequence of the invention as defined above. A particularly interesting polypeptide of the invention is a recombinant human β -casein polypeptide comprising the amino acid sequence shown in Figure 1 or a subsequence thereof having the calcium-binding activity of human β -casein. An example of an important
30 polypeptide comprising an important subsequence of said amino acid sequence is a polypeptide comprising amino acid residues 1-210 of the amino acid sequence shown in Figure 1 corresponding to the mature recombinant human β -casein without a signal peptide.

As it will be apparent from the above disclosure, another interesting polypeptide of the present invention is one which differs from a polypeptide comprising the amino acid sequence shown in Figure 1 in that at least one amino acid residue has been substituted with a different amino acid residue and/or in that at least one amino acid residue has been deleted or added so as to result in a polypeptide comprising an amino acid sequence being different from the amino acid sequence shown in Figure 1 and having a similar, increased or decreased calcium-binding activity as compared to the calcium-binding activity of human β -casein or having ACE-inhibitory activity or opioid activity. Examples of a strategy for designing and preparing modified polypeptides of the invention are apparent from the above disclosure.

The polypeptide of the present invention is one in which at least one amino acid residue has been modified by post-translational modification and is preferably in glycosylated and/or phosphorylated form. Normally, glycosylation is achieved when the polypeptide is expressed by a cell of a higher organism such as yeast or preferably a mammal, whereas phosphorylation is achieved also in the case of expression in lower organisms such as bacteria. The amino acid residues of the polypeptide of the invention which are normally phosphorylated are mentioned above. Glycosylation is normally found in connection with amino acid residues Asn, Ser, Thr or hydroxylysine.

In a further aspect, the present invention relates to a replicable expression vector which carries and is capable of mediating the expression of a DNA sequence encoding human β -casein.

In the present context, the term "replicable" means that the vector is able to replicate in a given type of host cell into which it has been introduced. Immediately upstream of the human β -casein DNA sequence there may be provided a sequence coding for a signal peptide, the presence of which ensures secretion of the human β -casein expressed by host cells harbouring the vector. The signal sequence may be the one naturally associated with the human β -casein DNA sequence or of another origin.

The vector may be any vector which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication; examples of such a vector are a plasmid, phage, cosmid, mini-chromosome or virus. Alternatively, the vector may be one which, when introduced in a host cell, is integrated in the host cell genome and replicated together with the chromosome(s) into which it has been integrated. Examples of suitable vectors are a bacterial expression vector, e.g. as exemplified in Example 2, and a yeast expression vector, e.g. as exemplified in Example 3. The vector of the invention may carry any of the DNA sequences of the invention as defined above and be used for the expression of any of the polypeptides of the invention defined above. However, when a bacterial vector is concerned it is preferred that the DNA sequence encoding human β -casein comprises the sequence shown in Sequence Listing 2, i.e. the sequence specifically adapted to expression in bacterial cells.

The present invention further relates to a cell harbouring a replicable expression vector as defined above. In principle, this cell may be of any type of cell, i.e. a prokaryotic cell such as a bacterium, e.g. *E. coli*, a unicellular eukaryotic organism, a fungus or yeast, e.g. *Saccharomyces cerevisiae* or a cell derived from a multicellular organism, e.g. a mammal. The mammalian cells are especially suitable for the purpose and are further discussed below.

In another important aspect, the invention relates to a method of producing recombinant human β -casein, in which a DNA sequence encoding human β -casein is inserted in a vector which is able to replicate in a specific host cell, the resulting recombinant vector is introduced into a host cell which is grown in or on an appropriate culture medium under appropriate conditions for expression of human β -casein and the human β -casein is recovered. The medium used to grow the cells may be any conventional medium suitable for the purpose. A suitable vector may be any of the vectors described above, and an appropriate host cell may be any of the cell types listed above. The

methods employed to construct the vector and effect introduction thereof into the host cell may be any methods known for such purposes within the field of recombinant DNA, examples of which are given in Example 2 and 3 below. The recombinant human β -casein expressed by the cells may be secreted, i.e. exported through the cell membrane, dependent on the type of cell and the composition of the vector. Examples of bacterial and yeast expression systems providing intracellular and extracellular expression, respectively, are given in Examples 2 and 3. The method outlined above is equally useful for the production of any of the polypeptides of the invention as defined above, i.e. on the basis of a DNA sequence of the invention.

If the human β -casein is produced intracellularly by the recombinant host, that is, is not secreted by the cell, it may be recovered by standard procedures comprising cell disruption by mechanical means, e.g. sonication or homogenization, or by enzymatic or chemical means followed by purification. Such methods are further described below and examples of the recovery procedure are given in Example 5. The present invention provides a generally novel purification method for purifying casein proteins including β -casein. This method is further discussed below.

In order to be secreted, the DNA sequence encoding human β -casein should be preceded by a sequence coding for a signal peptide, the presence of which ensures secretion of human β -casein from the cells so that at least a significant proportion of the human β -casein expressed is secreted into the culture medium and recovered.

Although recombinant production of human β -casein as disclosed above using lower organisms such as bacteria and yeast as production organisms for some purposes is satisfactory, e.g. when moderate yields of human β -casein are sufficient, when a short-term production is desirable or when human β -casein of a high purity substantially free from other mammalian derived substances such as proteins, in particular milk proteins, are desirable, the presently preferred method of producing recombinant human β -casein of the invention is by use of transgenic non-human mammals capable of excreting the human β -casein into their milk. The use of transgenic non-human mammals has the

advantage that large yields of recombinant human β -casein are obtainable at reasonable costs and, especially when the non-human mammal is a cow, that the recombinant human β -casein is produced in milk which is the normal constituent of, e.g., infant formulae so that no extensive purification is needed when the recombinant human β -casein is to be used as a nutrient supplement in milk-based products. Furthermore, production in a higher organism such as a non-human mammal normally leads to the correct processing of the mammalian protein, e.g. with respect to post-translational processing as discussed above and proper folding. Also large quantities of substantially pure human β -casein may be obtained.

Accordingly, in a further important aspect, the present invention relates to a mammalian expression system comprising a DNA sequence encoding human β -casein inserted into a gene encoding a milk protein of a non-human mammal so as to form a hybrid gene which is expressible in the mammary gland of an adult female of a mammal harbouring said hybrid gene.

The DNA sequence encoding human β -casein is preferably a DNA sequence as defined above encoding a polypeptide comprising the amino acid sequence shown in Figure 1 or a human β -casein gene or an analogue or effective subsequence thereof.

The mammary gland as a tissue of expression and genes encoding milk proteins are generally considered to be particularly suitable for use in the production of heterologous proteins in transgenic non-human mammals as milk proteins are naturally produced at high expression levels in the mammary gland. Also, milk is readily collected and available in large quantities. In the present connection the use of milk protein genes in the production of recombinant human β -casein has the further advantage that it is produced under conditions similar to the its natural production conditions in terms of regulation of expression and production location (the mammary gland).

In the present context the term "hybrid gene" denotes a DNA sequence comprising on the one hand a DNA sequence encoding human β -casein as defined above and on the other hand a DNA sequence of the milk prote-

in gene which is capable of mediating the expression of the hybrid gene product. The term "gene encoding a milk protein" denotes an entire gene as well as an effective subsequence thereof capable of mediating and targeting the expression of the hybrid gene to the tissue of interest, i.e. the mammary gland. Normally, the effective subsequence is one which at least harbours one or more of a promoter region, a transcriptional start site, 3' and 5' non-coding regions and structural sequences. The DNA sequence encoding human β -casein is preferably substantially free from prokaryotic sequences, such as vector sequences, which may be associated with the DNA sequence after, e.g., cloning thereof.

The hybrid gene is preferably formed by inserting *in vitro* the DNA sequence encoding human β -casein into the milk protein gene by use of techniques known in the art. Alternatively, the DNA sequence encoding human β -casein can be inserted *in vivo* by homologous recombination.

Normally, the DNA sequence encoding human β -casein will be inserted in one of the first exons of the milk protein gene of choice or an effective subsequence thereof comprising the first exons and preferably a substantial part of the 5' flanking sequence which is believed to be of regulatory importance.

The hybrid gene preferably comprises a sequence encoding a signal peptide so as to enable the hybrid gene product to be secreted correctly into the mammary gland. The signal peptide will typically be the one normally found in the milk protein gene in question or one associated with the DNA sequence encoding human β -casein. However, also other signal sequences capable of mediating the secretion of the hybrid gene product to the mammary gland are relevant. Of course, the various elements of the hybrid gene should be fused in such a manner as to allow for correct expression and processing of the gene product. Thus, normally the DNA sequence encoding the signal peptide of choice should be precisely fused to the N-terminal part of the DNA sequence encoding human β -casein. In the hybrid gene, the DNA sequence encoding human β -casein will normally comprise its stop codon, but not its own message cleavage and polyadenylation site. Downstream of

the DNA sequence encoding human β -casein, the mRNA processing sequences of the milk protein gene will normally be retained.

A number of factors are contemplated to be responsible for the actual expression level of a particular hybrid gene. The capability of the promoter as well of other regulatory sequences as mentioned above, the integration site of the expression system in the genome of the mammal, the integration site of the DNA sequence encoding human β -casein in the milk protein encoding gene, elements conferring post-transcriptional regulation and other similar factors may be of vital importance for the expression level obtained. On the basis of the knowledge of the various factors influencing the expression level of the hybrid gene, the person skilled in the art would know how to design an expression system useful for the present purpose.

A variety of different milk proteins are secreted by the mammary gland. Two main groups of milk proteins exist, namely the caseins and the whey proteins. The composition of milk from different species varies qualitatively as well as quantitatively with respect to these proteins. Most non-human mammals produces 3 different types of casein, namely α -casein, β -casein and κ -casein. The most common bovine whey proteins are α -lactalbumin and β -lactalbumin. The composition of milk of various origins are further disclosed in Clark et al., 1987.

The milk protein gene to be used may be derived from the same species as the one in which the expression system is to be inserted, or it may be derived from another species. In this connection it has been shown that the regulatory elements that target gene expression to the mammary gland are functional across species boundaries (which may be due to a possible common ancestor) (Hennighausen et al.).

Examples of suitable genes encoding a milk protein or effective subsequences thereof to be used in the construction of an expression system of the invention are normally found among whey proteins of various mammalian origins, e.g. a whey acidic protein (WAP) gene, preferably of murine origin, and a β -lactoglobulin gene, preferably of ovine origin. Also casein genes of various origins may be found to be suitable for the transgenic production of human β -casein, e.g.

bovine α S1-casein and rabbit β -casein. The presently preferred gene is a murine WAP gene as this has been found to be capable of providing a high level expression of a number of foreign human proteins in milk of different transgenic animals (Hennighausen et al.).

- 5 Another sequence preferably associated with the expression system of the invention is a so-called expression stabilizing sequence capable of mediating high-level expression. Strong indications exist that such stabilizing sequences are found in the vicinity of and upstream of milk protein genes.
- 10 The DNA sequence encoding a human β -casein to be inserted in the expression system of the invention may be of cDNA, genomic or synthetic origin or any combination thereof. While some expression systems have been found to function best when cDNA encoding a desirable protein is used, others have been found to require the presence of
- 15 introns and other regulatory regions in order to obtain a satisfactory expression (Hennighausen et al.). In some cases it may be advantageous to introduce genomic structures as polypeptide encoding element in vector constructs compared to cDNA elements (Brinster et al.). The intron and exon structure may result in higher steady state
- 20 mRNA levels that obtained when cDNA based vectors are used. When the expression system is based on the WAP gene, a cDNA sequence is preferred as it has been verified by various experiments that the WAP gene is capable of providing as good an expression of cDNA as of genomic DNA, or even a better expression (Hennighausen et al.). In
- 25 Example 4, an expression system based on the murine WAP gene and a cDNA sequence of the invention encoding human β -casein is illustrated and further discussed.

- 30 In a further aspect, the present invention relates to a hybrid gene comprising a DNA sequence encoding human β -casein inserted into a gene encoding a milk protein of a non-human mammal, the DNA sequence being inserted in the milk protein gene in such a manner that it is expressible in the mammary gland of an adult female of a mammal
- 35 harbouring the hybrid gene. The hybrid gene and its constituents have been discussed in detail above. The hybrid gene constitutes an impor-

tant intermediate in the construction of an expression system of the invention as disclosed above.

In another aspect, the present invention relates to a non-human mammalian cell harbouring an expression system as defined above. The mammalian cell is preferably an embryo cell or a pro-nucleus. The expression system is suitably inserted in the mammalian cell using a method as explained in the following and specifically illustrated in Example 4 below.

In a further important aspect, the present invention relates to a method of producing a transgenic non-human mammal capable of expressing human β -casein, comprising injecting an expression system of the invention as defined above into a fertilized egg or a cell of an embryo of a mammal so as to incorporate the expression system into the germline of the mammal and developing the resulting injected fertilized egg or embryo into an adult female mammal.

The incorporation of the expression system into the germline of the mammal may be performed using any suitable technique, e.g. as described in "Manipulating the Mouse Embryo"; A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1986. For instance, a few hundred molecules of the expression system may be directly injected into a fertilized egg, e.g. a fertilized one cell egg or a pro-nucleus thereof, or an embryo of the mammal of choice and the microinjected eggs may then subsequently be transferred into the oviducts of pseudopregnant foster mothers and allowed to develop. Normally, not all of the injected eggs will develop into adult females expressing human β -casein. Thus, about half of the mammals will from a statistically point of view be males from which, however, females can be bred in the following generations.

Once integrated in the germ line, the DNA sequence encoding human β -casein may be expressed at high levels to produce a correctly processed and functional human β -casein in stable lines of the mammal in question.

Of further interest is a method of producing a transgenic non-human mammal capable of expressing human β -casein and substantially incapable of expressing β -casein from the mammal itself, comprising

- 1) destroying the mammalian β -casein expressing capability of the mammal so that substantially no mammalian β -casein is expressed and inserting an expression system of the invention as defined above or a DNA sequence encoding human β -casein into the germline of the mammal in such a manner that human β -casein is expressed in the mammal and/or
- 2) replacing the mammalian β -casein gene or part thereof with an expression system of the invention as defined above or a DNA sequence encoding human β -casein.

The mammalian β -casein expressing capability is conveniently destroyed by introduction of mutations in the DNA sequence responsible for the expression of the β -casein. Such mutations may comprise mutations which make the DNA sequence out of frame, or introduction of a stop codon or a deletion of one or more nucleotides of the DNA sequence.

The mammalian β -casein gene or a part thereof may be replaced with an expression system as defined above or a DNA sequence encoding human β -casein by use of the well known principles of homologous recombination.

In a further aspect, the present invention relates to a transgenic non-human mammal prepared by a method as described above.

While the transgenic non-human mammal of the invention in its broadest aspect is not restricted to any particular type of mammal, the mammal will normally be selected from the group consisting of mice, rats, rabbits, sheep, pigs, goats and cattle. For large scale production of human β -casein the larger animals such as sheep, goats, pigs and especially cattle are normally preferred due to their high milk production. However, also mice, rabbits and rats may be interesting due to the fact that the manipulation of these animals is more simple

and results in transgenic animals more quickly than when, e.g. cattle, are concerned.

Also progeny of a transgenic mammal as defined above, capable of producing human β -casein is within the scope of the present invention.

In a still further aspect, the present invention relates to a method of preparing human β -casein comprising collecting milk from a transgenic non-human mammal of the invention as defined above and recovering the human β -casein from the milk. The milk may be collected in any suitable manner normally used in connection with the collection of milk from the mammal in question.

From the above explanation it will be clear that the present invention for the first time makes it possible to produce milk from a non-human mammal comprising human β -casein, the importance and utility of which will be apparent from the present context. Thus, a further aspect of the present invention includes milk from a non-human mammal comprising recombinant human β -casein. Of particular interest is milk from a non-human mammal comprising a polypeptide of the invention as defined above comprising the amino acid sequence shown in Figure 1 or a polypeptide encoded by a DNA sequence or an analogue or subsequence thereof as defined above. Typically, the milk of the invention will be obtained from a transgenic mammal of the invention as defined above.

From the above explanation it will be apparent that an important use of the polypeptide of the invention is as a nutrient supplement, in particular as a substituent of an infant formula.

In a still further aspect, the present invention relates to an infant formula comprising recombinant human β -casein, in particular a polypeptide of the invention as defined above. The infant formula may be prepared by adding the recombinant human β -casein or polypeptide in a purified or partly purified form to the normal constituents of the infant formula. However, normally it is preferred that the infant formula is prepared from milk of the invention as defined above,

especially when it is of bovine origin. The infant formula may be prepared using conventional procedures and contain any necessary additives such as minerals, vitamins etc.

5 In a further aspect, the present invention relates to a generally novel method of isolating and/or purifying casein proteins from a mixture of components including proteins different from casein. The method comprises adding to the mixture of components from which the casein is to be isolated and/or purified ammonium sulphate in a concentration equal to or of above 0.05 M to precipitate the casein
10 proteins, separating the precipitated β -casein protein from the mixture of components, and optionally removing salt from the precipitated casein.

The purification method of the invention is based on the fact that proteins show a decreased solubility in the presence of salts. This
15 effect is known as "salting out".

The method is believed to be generally applicable for the isolation of any kind of casein proteins from any kind of mixture, including a mixture in which the casein proteins are naturally present, e.g. milk, and a mixture in which the casein proteins have been artificially added or synthesized, e.g. from a mixture of cells or cell
20 components in which recombinant casein has been produced. The cells may be microorganism cells, such as bacterial or yeast cells, or mammalian cells.

Conventionally, casein in milk has been separated by acid precipitation at its isoelectrical point, e.g. as described by Rowland, 1938.
25 However, acid precipitation has certain disadvantages including the need of pH-control, which disadvantages are especially pronounced in connection with large-scale casein purification.

Thus it is surprising to find that ammonium sulphate in a concentration as low as 0.05 M in the separation of casein from the mixture of
30 components, e.g. as described in Example 5, provided a very selective casein separation method as compared to conventional methods. The use of ammonium sulphate at the low concentration is inexpensive and may

easily be used in large-scale isolation and/or purification of casein.

Although the various types of caseins, e.g. α -casein, β -casein, and κ -casein, cannot be separated from each other by use of the method disclosed above, it is believed that on the basis of the very selective casein separation method of the invention, the various types of caseins optionally present in this isolated casein precipitate can be further separated, if necessary by chromatography.

More particularly, the casein isolating method of the present invention may be used in the isolation of recombinant human β -casein from a mixture of bacterial or yeast cells in which it is produced, e.g. as disclosed above. Thus, in a further aspect, the present invention relates to a method of isolating recombinant human β -casein from bacterial or yeast cells as defined above which, in the case of intra-cellular expression of recombinant human β -casein in the bacterial or yeast cells, comprises separating the cells harboring the recombinant human β -casein from the culture medium, rupturing the separated cells so as to allow them to release their content of *inter alia* recombinant human β -casein, optionally removing cell debris from the mixture of ruptured cells and precipitating the released β -casein by addition of ammonium sulphate at a concentration equal to or above 0.05 M.

The precipitate is consequently separated from the mixture in which it is present, the salt is optionally removed and the β -casein is concentrated, e.g. by freeze drying.

The separation of the bacterial or yeast cell from the culture medium, the optional removal of the cell debris, and the separation of the precipitate are conveniently carried out by centrifugation or by sedimentation or by other conventionally used separation methods. The rupturing of the bacterial or yeast cells may be accomplished by any suitable method such as by sonication, rupture by a freeze/thaw treatment, a French Press treatment or the like.

Any salt may be removed by conventional methods, e.g. by dialysis or size-exclusion chromatography. The isolated and optionally purified β -casein is preferably concentrated by ultrafiltration and/or lyophilization.

- 5 The β -casein precipitate may advantageously be subjected to one or more wash treatments before it is concentrated, or, in the event that the method comprises removal of salt, before this is done.

When the recombinant human β -casein has been produced extra-cellularly, it is in most cases not necessary to rupture the bacterial or yeast cells because the recombinant human β -casein has already been excreted into the medium. However, if the extra-cellular expression is not satisfactorily effective and a large part of the produced recombinant human β -casein is present in the cells, it may of course be an advantage to rupture such cells. However, normally when extra-cellular expression is concerned, the method of isolating recombinant human β -casein from bacterial or yeast cells as defined above comprises removing the bacterial or yeast cells from the culture medium, precipitating the recombinant human β -casein present in the culture medium by addition of ammonium sulphate in a concentration equal to or above 0.05M, separating the precipitated recombinant human β -casein from the medium, and optionally removing the salt and subjecting the isolated β -casein to concentration.

When recombinant human β -casein of the invention is to be separated from milk produced by a transgenic non-human mammal as described above, the ammonium sulphate is added directly to the milk and the precipitated casein comprising the recombinant human casein is recovered from the milk. The casein fraction which may comprise other types of casein than the recombinant human β -casein, e.g. bovine casein proteins, may be subjected to further purification so as to purify, e.g., the recombinant human β -casein therefrom, e.g. by chromatography or fractionated desalting. The ammonium sulphate concentration used in the purification method of the invention is preferably above 0.05 M, such as equal to or above 0.06 M, e.g. equal to or above 0.07 M. However, also higher concentrations such as con-

centrations equal to or above 0.075 M, 0.08 M and even 0.09 and 0.1 M are contemplated to be suitable for the purpose.

The invention is further described with reference to the accompanying drawings in which

5 Figure 1 discloses the complete nucleotide and deduced amino acid sequences of the human β -casein cDNA fragment obtained as described in Example 1. The position of the oligonucleotide probe used for screening of the cDNA library is shown by underlining. The broken arrows show the positions of the different oligodeoxyribonucleotide
10 primers used for sequencing of the cDNA fragment,

Figure 2 the human β -casein cDNA sequence useful for expression in *E. coli* and cloned as an EcoRI fragment into pUC 19 (as described in Example 2). The cDNA was modified by introduction of synthetic oligo-
15 nucleotides in both the 5' and 3' termini. The modifications altered the codon usage in the 5' end, added an extra stop codon to the 3' end and introduced restriction sites which facilitate cloning into the expression vectors,

Figure 3 the expression vector pS 26 constructed as described in Example 2 and useful for intra-cellular expression of human β -casein
20 in *E. coli*,

Figure 4 the expression vector pS 28 constructed as described in Example 2 and useful for extra-cellular expression of human β -casein in *E. coli*,

Figure 5 the result of an SDS-PAGE analysis of soluble β -casein expressed in *E. coli* BL21 (DE3) pLys S performed as described in the
25 legend to Figure 7.

A) SDS-PAGE separation of soluble intracellular proteins of IPTG induced *E. coli* BL21 (DE3) pLys S carrying pS 26. The gel was stained with Coomassie Brilliant Blue R-250.
30 Lane 1: Molecular weight standard 94, 67, 43, 30, 20.2 and 14 kD (Pharmacia).

Lanes 2-5. Soluble protein fraction 1, 2, 3 and 8 μ l, respectively.

B) Western blot of the same fraction as in A.

Lanes 1-4 contain the same samples as lanes 2-5 in A.

5 Lane 5: Prestained molecular weight standard 110, 84, 47, 33, 24 and 16kD (Bio-Rad). The samples were separated by SDS-PAGE, transferred to Immobiline (Millipore) membranes, visualized using polyclonal rabbit antibodies produced using highly purified human β -casein, followed by an alkaline phosphatase labelled swine anti-rabbit Ig (Dakopatts).

10 C) SDS-PAGE separation of soluble protein fraction. The gel was stained as in A.

Lane 1: *E. coli* BL21 (DE3) pLys S uninduced.

Lane 2: As 1, but IPTG induced.

Lane 3: *E. coli* BL21 (DE3) pLys S transformed with pS 28 uninduced.

15 Lane 4: As lane 3, but IPTG induced.

Lane 5: Molecular weight standard as in A, Lane 1.

Lane 6: Purified human native β -casein,

Figure 6 the expression vector pS 232 used for extra-cellular expression of human β -casein in *S. cerevisiae* and further described in
20 Example 6.

Figure 7 the result of an Western blot analysis of cell extracts and growth medium from yeast cells containing the pS232 plasmid. The proteins are separated on a 10-17.5% gradient polyacrylamide gel followed by electroblotting to a nitrocellulose filter. Prior to
25 hybridization the filter was blocked with 2.5% BSA. As a first antibody purified rabbit polyclonal antibodies against β -casein were used and as a second antibody Dakopatts alkaline phosphatase conjugated affinity purified swine immunoglobulines to rabbit IgG were used. The filter was developed with nitro blue tetrazolium and 5-
30 bromo-4-chloro-3-indolylphosphatase.

Lane 1: Molecular weight marker 110, 84, 47, 24 and 16kD.

Lane 2: pS 232/UMY 519 25 μ l cell lysate.

Lanes 3 and 4: blank.

Lane 5: pS 232/UMY 519 25 μ l culture medium.

- Lane 6: 3 μ g of purified native human β -casein
Lane 7: Purified recombinant human β -casein produced in *E. coli* BL21 (DE3) pLys S containing pS 26.
Lane 8: pYES 2.0/UMY 519 25 μ l cell lysate.
5 Lane 9: Blank.
Lane 10: pYES 2.0/UMY 519 25 μ l culture medium.

Figure 8

A) the organization of the human β -casein cDNA inserted into the mouse WAP gene as described in Example 4,

- 10 B) the expression vector pS 133 used for expression of human β -casein in mice and constructed as described in Example 4,

Figure 9 the sequence of the KpnI and SalI fragment generated by PCR as described in Example 4,

- Figure 10 the result of a western blot analysis of milk samples
15 derived from three mouse lines transgenic for human β -casein, carrying the pS133 derived vector fragment (Example 4). The proteins are separated on SDS-PAGE transferred to Immobilon membranes (Millipore) and visualized by polyclonal rabbit antibodies produced using highly purified human β -casein, followed by alkaline phosphatase labelled
20 swine anti-rabbit Ig (Dakopatts). In each of the lanes, 10 μ l of the whey fraction was loaded.

- Lane 1: pS 133 transgene line 69.
Lane 2: pS 133 transgene line 72.
Lane 3: pS 133 transgene line 79.
25 Lane 4: Non-transgene control line.
Lane 5: 50 ng of recombinant *E. coli* produced human β -casein.

Figure 11 the results obtained after purification of recombinant human β -casein as described in Example 5.

- A) SDS-PAGE analysis of purified recombinant β -casein produced in *E. coli*. The gel was stained with Coomassie Brilliant Blue R-250.
30 Lane 1: Low molecular weight standard (Pharmacia).

Lanes 2-4: Ammoniumsulphate precipitated β -casein at increasing concentrations.

B) Western blot analysis of recombinant β -casein purified as in A).

Lanes 1 and 5: Prestained low molecular weight standard (BioRad).

5 Lane 2: Native β -casein purified from human milk.

Lanes 3 and 4: Recombinant β -casein purified as in A), and

10 Figure 12 the purity of recombinant human β -casein purified as described in Example 5 and assayed by ion-exchange chromatography. The sample was loaded onto the column in 20mM ethanolamine-HCl, 6M urea, pH 9.5 and eluted with a linear gradient up to 0.6M NaCl in the same buffer.

References

- Brantl, V. Novel opioid peptides derived from human β -casein: Human δ -casomorphins. *Eur. J. Pharmacol.* 106, 213-214, 1984.
- Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and
5 Palmiter, R.D. Introns increase transcriptional efficiency in trans-
genic mice. *Proc. Natl. Acad. Sci. USA* 85, 836-840, 1988.
- Campbell, S.M., Rosen, J.M., Hennighausen, L.G., Strech-Jurk, U. and
Sippel, A.E. Comparison of the whey acidic protein genes of the rat
and mouse. *Nucleic Acids Res.* 12, 8685-8697, 1984.
- 10 Clarl, A.J., Simons, P., Wilmut, I. and Lathe, R. Pharmaceuticals
from transgenic livestock. *TIBTECH* 5, 20-24, 1987.
- Gill, D.R., Hatfull, G.F. and Salmond, G.P.C. A new cell division
operon in *E. coli*. *Mol. Gen. Genet.* 205, 134-145, 1986.
- Greenberg, R., Groves, M.L. and Down, H.J. Human β -casein, amino acid
15 sequence and identification of phosphorylation sites. *J. Biol. Chem.*
259, 5132-5138, 1984.
- Hennighausen, L.G., Ruiz, L. and Wall, R. Transgenic animals - pro-
duction of foreign proteins in milk. *Curr. Opinions Biotechn.* 1, 74-
78, 1990.
- 20 Jiminez-Flores, R., Kang, Y.C. and Richardson, T. *Biochem. Biophys.*
Res. Commun. 142, 617-621, 1987.
- Kohmura, M., Nio, N., Kubo, K., Minoshima, Y., Munekata, E. and Ari-
yoshi, Y. Inhibition of Angiotensin-converting enzyme by synthetic
peptides of human β -casein. *Agric. Biol. Chem.* 53, 2107-2114, 1989.
- 25 Kunz, C. and Lønnerdal, B. Human milk proteins: analysis of casein
and casein subunits by anion-exchange chromatography, gel electropho-
resis, and specific staining methods. *Am. J. Clin. Nutr.* 51, 37-46,
1990.

- Laemmli, U.K. *Nature* 227, 680, 1970.
- Lönnerdal, B., Bergström, S., Andersson, Y., Hjalmarsson, K., Sundqvist, A.K. and Hernell, O. Cloning and sequencing of a cDNA encoding human milk β -casein. *FEBS Lett.* 269, 153-156, 1990.
- 5 Maruyama, S., Nakagomi, K., Tomizuka, N. and Suzuki, H. Angiotensin I-converting enzyme inhibitor derived from an enzymatic hydrolysate of casein. Isolation and bradykinin-potentiating activity on the uterus and the ileum of rats. *Agric. Biol. Chem.* 49, 1405-1410, 1985.
- 10 Menon, R.S. and Ham, R.G. Human β -casein: partial cDNA sequence and apparent polymorphism. *Nucleic Acids Res.* 17, 2869, 1989.
- Picken, R.N., Mazaitis, A.J., Maas, W.K., Rey, M. and Heyneker, H. Nucleotide sequence of the gene for heat-stable enterotoxin II of *E. coli*. *Infect. Immun.* 42, 269-275, 1983.
- 15 Provot, C., Persuy, M.A. and Mercier, J.C. Complete nucleotide sequence of ovine beta-casein cDNA: inter-species comparison. *Biochimie* 71, 827-32, 1989.
- Rowland, S.J. *J. Dairy Res.* 9, 47-57, 1938.
- Sambrook, J., Fritsch, E.F. and Maniatis, T.E. *Molecular cloning, a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory Press, 1989.
- 20 Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. Use of T7 RNA Polymerase to direct expression of cloned genes. In *Methods in Enzymology*, ed. David V. Goeddel, p. 60-89. Academic Press, 1990.
- Miller, Mark J. S. et al., *Proc. Soc. Exp. Biol. Med.* 195, 143-159, 1990.

Additional references dealing with cDNA and genes coding for milk proteins.

11

5 Bonsing, J. and Mackinlay, A.G. Recent studies on nucleotide sequences encoding the caseins. J. Dairy Res. 54, 447-61, 1987.

12

Hall, L., Emery, D.C., Davies, M.S., Parker, D. and Craig, R.K. Organization and sequence of the human alpha-lactalbumin gene. Biochem. J. 242, 735-42, 1987.

10 13

Blackburn, D.E., Hobbs, A.A. and Rosen, J.M. Rat beta casein cDNA: sequence analysis and evolutionary comparisons. Nucleic Acids Res. 10, 2295-2307, 1982.

14

15 Jones, W.K., Yu-Lee, L.Y., Clift, S.M., Brown, T.L. and Rosen, J.M. The rat casein multigene family. Fine structure and evolution of the beta-casein gene. J. Biol. Chem. 260, 7042-7050, 1985.

15

20 Yoshimura, M., Banerjee, M.R. and Oka, T. Nucleotide sequence of a cDNA encoding mouse beta casein. Nucleic Acids Res. 14, 8224, 1986.

16

Yoshimura, M. and Oka, T. Isolation and structural analysis of the mouse beta-casein gene. Gene 78, 267-275, 1989.

17

25 Menon, R.S and Ham, R.G. Human β -casein: partial cDNA sequence and apparent polymorphism. Nucleic Acids Res. 17, 2869, 1989.

18

Stewart, A.F., Bonsing, J., Beattie, C.W., Shah, F., Willis, I.M., Mackinlay, A.G. Complete nucleotide sequences of bovine alpha S2- and

beta-casein cDNAs: comparisons with related sequences in other species. Mol. Biol. Evol. 4, 231-41, 1987.

19

Provot, C., Persuy, M.A. and Mercier, J.C. Complete nucleotide sequence of ovine beta-casein cDNA: inter-species comparison. Biochimie 71, 827-32, 1989.

20

Devinoy, E., Schaerer, E., Jolivet, G., Fontaine, M.L., Kraehenbuhl, J.P. and Houdebine, L.M. Sequence of the rabbit alpha S1-casein cDNA. Nucleic Acids Res. 16, 11813, 1988.

21

Schaerer, E., Devinoy, E., Kraehenbuhl, J.P. and Houdebine, L.M. Sequence of the rabbit beta-casein cDNA: comparison with other casein cDNA sequences. Nucleic Acids Res. 16, 11814, 1988.

EXAMPLES

The following examples are intended to illustrate but not to limit the present invention.

5 Construction of the expression systems of the invention, and the molecular biological characterization of it, employs standard methods generally known in the art of recombinant DNA. Unless otherwise stated, the methods used are those described by Sambrook et al., 1989.

DEFINITIONS

10 Hybridization of DNA

DNA, e.g. present on nitrocellulose filters, are wetted in 2 x SSC [1 x SSC: 0.15 M NaCl, 0.0015 M Na₂-citrate, pH 7.0] and placed in a heat-sealed plastic bag with pre-warmed (67°C) prehybridization solution. Prehybridization takes place for 2 h at 67°C, the bag being
15 gently shaken. The solution is exchanged with pre-warmed (67°C) hybridization solution, the radioactive probe is added and hybridization is carried out at 67°C for 18 h. The bag is gently shaken to ensure constant movement of the liquid over the nitrocellulose filters. After hybridization, a washing procedure is carried out.

20 The radioactive probe is prepared by use of known methods, e.g. as described by Sambrook et al., on the basis of the DNA sequence shown in Sequence Listing 1 or a part thereof, especially a coding part such as the nucleotides corresponding to amino acids 1-210 or an effective subsequence of the DNA sequence as defined above.

25 The prehybridization and hybridization solutions used are: 10 x Denhardt's, 4 x SSC, 0.1% SDS, 10 µg/ml polyA, 50 µg/ml of denatured DNA to be analysed and the denatured (heat) radioactive probe. The filters are washed in pre-warmed (67°C) solutions: 10 x Denhardt, 2 x SSC, 0.1% SDS for 2 x 15 min. and 1 x SSC, 0.1% SDS for 4 x 15
30 min. The filters are air-dried and covered with Vita-Wrap, and X-ray

film is exposed to the filters for 3 h to 3 weeks with and without intensifying screens.

EXAMPLE 1A

Cloning and sequencing of cDNA encoding human β -casein

5 A λ -gt 11 human mammary gland cDNA library prepared from lactational-ly competent adult human mammary gland was obtained from Clontech Lab., Palo Alto, CA. The clones of the human β -casein were screened by plaque hybridization using *E. coli* Y 1090. A synthetic 42-mer oligonucleotide probe

10 5'-GAGCAAGGGAAGAGGCAAATGAAGATTTTCAAGATCAGTCAA-3'

corresponding to amino acids 117-130 in the β -casein sequence (Greenberg et al., 1984) was synthesized. The construction of this oligonucleotide was based on a region of the bovine β -casein amino acid sequence having a large degree of homology with the human counter-
15 part. However, the selection of the synthetic oligonucleotide included the following modifications of the bovine cDNA sequence (Jimenez-Flores et al., 1987) for the nucleotide sequence 449-490: i) nucleotide 458 (G \rightarrow C exchange), ii) nucleotide 480 (C \rightarrow T), iii) nucleotide 483 (T \rightarrow C). The oligonucleotide probe was synthesized on a Beck-
20 man 200A DNA synthesizer using the phosphoramidite technique according to the vendor's instructions. The probe was [γ - 32 P] dATP-labelled using T4 polynucleotide kinase (New England Biolabs; Beverly, MA). Hybridization was carried out for 12-15 hours at 40°C, and the membranes were washed and autoradiographed on X-ray film (Amersham, UK).
25 Six positive plaques were identified in the primary screening.

Following secondary screening, phage DNA of purified clones were isolated from plate lysates. Restriction mapping was performed and the β -casein cDNA was localized in the cloned fragment by Southern blotting.

One of the λ -gt 11 clones carrying an insert hybridizing to the β -casein 42-mer probe was digested with the restriction end nuclease EcoRI, and the cDNA insert was separated from DNA by electrophoresis in 1% Sea Kem GTG Agarose (FMC BioProducts, ME). The cDNA fragment was ligated to EcoRI-digested alkaline phosphatase-treated pUC19 DNA (Pharmacia) and transformed into *E. coli* TG1 (Studier et al.). Transformants were selected on plates containing 100 μ g/ml of carbenicillin, 40 μ g/ml of 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) and 1 mM isopropyl- β -D-thiogalactoside (IPTG, Sigma, St. Louis, MO). A recombinant plasmid carrying the cDNA insert was identified and designated pS 21. Plasmid pS 21 DNA was subjected to restriction endonuclease analysis. The complete nucleotide sequence of both strands of the region encoding β -casein was determined, using T7 sequencing kit (Pharmacia, Uppsala, Sweden), on double stranded templates as described by the vendor. As primers for sequencing reactions, specific oligonucleotides complementary to pUC19 or β -casein sequences were used.

The nucleotide sequence contained an open reading frame sufficient to encode the entire amino acid sequence of a β -casein precursor protein consisting of 210 amino acids and a signal peptide of 15 amino acids (Figure 1).

EXAMPLE 1B

Genetic variants of human β -casein

It is assumed that human β -casein exists in a limited number of genetic variants. These variants will have a number of amino acid substitutions as compared to the amino acid sequence deduced from the cDNA sequence shown in Figure 1. The assumption is based on the fact that most other species investigated to date do have genetic variants, but also on the discrepancies found between the obtained cDNA sequence (Figure 1) and the sequence determined by classical amino acid sequencing (Greenberg et al.). Genetic variants, i.e. analogues of the DNA sequence shown in Figure 1, may be isolated and characterized by the following procedure:

DNA is isolated from fresh human milk provided by donors with varying genetic background (ethnicity). Similarly, mRNA is isolated from fresh milk and cDNA is synthesized by use of the reverse transcriptase methodology. Using specific synthetic oligonucleotides, selected from regions flanking sequences with pronounced amino acid discrepancies, DNA fragments are synthesized by the use of the PCR technique. Synthesized DNA fragments are isolated from agarose gels and sequenced by the dideoxy chain-termination method.

EXAMPLE 1C

10 Cloning, sequencing and organization of genomic DNA coding for human β -casein

A genomic clone containing the transcribed part of the human β -casein gene is isolated using the cloned human β -casein described in Example 1A as a probe.

15 The probe is the 1075 bp EcoRI fragment derived from pS 21, isolated by agarose gel electrophoresis and radioactively labelled by using a multiprime DNA labelling system (Amersham) according to the supplier's instructions. This labelled probe is used for screening of a human genomic library cloned in a lambda vector. The screening procedure is as described in Sambrook et al., 1989.

The transcribed region of the human β -casein gene is expected to be in a 7 kb genomic fragment. The overall structure of the gene is expected to show more than four exons, probably nine exons, interrupted by introns located within a 10 kb, probably within the 7 kb genomic fragment. The cloned genomic fragment is sequenced by using the same procedures as described above.

EXAMPLE 2

Expression of recombinant human β -casein in *Escherichia coli*

The cDNA encoding the pro-polypeptide for human β -casein was isolated as described above as a 1075 bp EcoRI fragment and cloned into pUC19, generating pS 21.

The cDNA termini were modified as follows (Figure 2), pS 21 was digested with EcoRI and AccI and a 328 bp fragment was isolated by agarose gel electrophoresis. The isolated 328 bp fragment was purified from the agarose gel by electroelution and digested with AvaII, and an AvaII and AccI 197 bp fragment was isolated.

Oligonucleotides with suitable restriction sites and altered codon usage adapted for *E. coli* expression encoding the amino terminus of mature β -casein, SYM 1328, SYM 1329, SYM 1330, SYM 1331, SYM 1332 and SYM 1333, were synthesized. The sequences are listed below:

SYM 1328 5'-CTCGAGCGAAGAATCGATCACCGAA-3'
 SYM 1329 5'-GAATTCATATGCGTGAAACCATCGAATCCCTGAG-3'
 SYM 1330 5'-TACAAAAAAGTTGAAAAAGTTAAACACGAGGACCAG-3'
 SYM 1331 5'-CTTTTTTGTATTGCGTGATCGATTCTTCGCTCGAGCTCAGGGATT-3'
 SYM 1332 5'-GATCCTGGTCCTCGTGTTTAACTTTTTTCAA-3'
 SYM 1333 5'-CGATGGTTTTCACGCATATGAATTCTGCA-3'

The six oligonucleotides described above were annealed and ligated into PstI and BamHI cleaved pUC19, and the resulting plasmid was sequenced and designated pS 24. An 89 bp PstI and AvaII restriction fragment was isolated from pS 24.

For modification of the 3' end, pS 21 was digested with AccI and a 641 bp fragment was isolated; this fragment was electroeluted and redigested with BglII, and a 303 bp fragment was isolated. The 303 bp fragment was subcloned into AccI and BglII digested pS 2 (Symbicom), generating pS 22. pS 22 was digested with EcoRI and BglII, and a fragment of approximately 2.96 kb was isolated.

In order to introduce suitable restriction sites and modify the 3' untranslated sequence, oligonucleotides were synthesized and ligated with the 2.96 kb fragment derived from pS 22.

The following oligonucleotides were used:

- 5 SYM 1335 5'-CCAGTTCATAACCCCATTAGTGTCTAATAAGGATCCG-3'
 SYM 1336 5'-GATCTACCCTGTGACTCAGCCACTTGCC-3'
 SYM 1338 5'-AATTCGGATCCTTATTAGACACTAATGG-3'
 SYM 1339 5'-GGTTATGAACTGGGGCAAGTGGCTGAGTCACAGGGTA-3'

The resulting plasmid was sequenced and designated pS 23.

- 10 To obtain the modified fragment encoding the mature β -casein, the following three fragments were ligated: first a PstI and AccI fragment of approximately 3.0 kb derived from pS 23, second the PstI and AvaII 89 bp fragment derived from pS 24, and third the 197 bp AvaII and AccI fragment from pS 21, resulting in pS 25.
- 15 The β -casein encoding sequence without signal was isolated from pS 25 as a 641 bp NdeI and BamHI fragment and introduced into the NdeI and BamHI digested vector plasmid pS14, and the resulting β -casein expression vector was designated pS 26 (Figure 3).

- 20 The vector pS 26 carries the bacteriophage T7 ϕ 10 promoter and ϕ terminator (Studier, 1990) to regulate the expression of β -casein. It also contains the origin of replication and sequences encoding ampicillin and resistance of pBR322. This vector was analyzed with appropriate enzymes and relevant segments were sequenced.

- 25 In order to allow secretion of recombinant human β -casein, a sequence encoding the signal peptide of heat stable enterotoxin II, STII, of *Escherichia coli* was introduced.

- 30 pS 25 was digested with Aval and EcoRI and a 619 bp fragment encoding the major part, except for the amino terminus, of β -casein was isolated. This fragment was cloned into NdeI and EcoRI cleaved pUC19 together with synthetic olig nucleotides. The synthetic oligonucleo-

tides encode the very amino terminal end of mature β -casein with appropriate restriction sites allowing fusion with the STII signal sequence. The following oligonucleotides were used:

SYM 1495 5'-TATGCACGTGAAACCATCGAATCCCTGAGC-3'
 5 SYM 1500 5'-TCGAGCTCAGGGATTCGATGGTTTCACGTGCA-3'

This three fragment ligation resulted in the plasmid pS 27. This plasmid was confirmed by sequence analysis.

In the next step pS 27 was cleaved with NdeI and HindIII and a 700 bp fragment was isolated and sequenced. This 700 bp fragment was introduced into the STII containing expression vector pS 29 which was
 10 digested with HindIII and NdeI. The resulting expression vector for human recombinant β -casein was designated pS 28 (Figure 4). pS 28 contains the same regulatory elements, replication signals and resistance markers as pS 26. The construction was confirmed by restriction
 15 analysis.

The expression vectors pS 26 and pS 28 were transformed into the following *Escherichia coli* strains: BL21(DE3), BL21(DE3) pLysE and BL21(DE3) pLysS (Studier, 1990). The expression experiments were carried out essentially as described by Studier et al., 1990. The
 20 obtained results demonstrated that recombinant human β -casein was efficiently expressed as a soluble protein with the two different expression vectors pS 26 and pS 28 (Figure 5). However, the *E. coli* produced human β -casein shows a somewhat lower apparent molecular weight.

25 EXAMPLE 3

Expression of recombinant human β -casein in *Saccharomyces cerevisiae*

A fragment encoding the entire β -casein polypeptide was isolated as a 698 bp KpnI and SalI fragment from pS133 (Figure 8, and further described in Example 4) by using agarose gel electrophoresis. The
 30 isolated fragment was ligated with KpnI and XhoI digested pYES2.0 (In

Vitrogen Corporation, USA). The resulting β -casein expression vector, designated pS 232 (Figure 6), was analyzed by restriction enzyme mapping.

The expression vector contains a polylinker and is designed for high
5 level expression in *S. cerevisiae*. The vector is designed for maintenance in both *E. coli* and yeast.

The vector contains the Gal 1 portion of the Gal 1/Gal 10 promoter region from *S. cerevisiae* for inducible expression of genes inserted into the polylinker. It also contains the CYC1 transcription termination signal, the 2 μ origin of replication and partitioning element
10 and the URA 3 gene for selection in yeast. The ampicillin resistance gene and the replication origin derived from pUC 19 as well as the T7 RNA polymerase promoter and F1 origin are present for maintenance and selection, transcription and single strand rescue, respectively, in
15 *E. coli*.

The Gal 1 promoter contains the sequences required for the regulation of transcription initiation as well as the mRNA start site, but does not contain an ATG for translation initiation. Therefore, the translation initiation must be located at the inserted sequence.

20 The expression vector pS 232 was transformed into the *Saccharomyces cerevisiae* host strain UMY 519 (URA⁻, Gal⁺) (A. Byström, personal communication), and selection was performed in uracil deficient medium. Colonies growing at restrictive conditions were isolated and analyzed.

25 A positive colony was cultured in minimal medium lacking uracil and containing 2% glucose to a density of about 0.5 OD₆₀₀. The culture was then centrifuged and the pellet was washed with distilled water. The washed pellet was divided. One part was resuspended in the same medium as above, and the other part was resuspended in the same
30 medium except that glucose was substituted with 2% galactose. The two pellets were resuspended to a density of about 0.1 OD₆₀₀. The two cultures were grown overnight to a density of about 3.0 OD₆₀₀, and the cultures were harvested.

After centrifugation of the cultures, the pellets were resuspended and the supernatants were mixed with SDS, urea and mercaptoethanol containing sample buffer (Laemmli et al., 1970) and boiled for 20 minutes. The samples were then loaded on a SDS-PAGE 10-17.5% gradient gel (Laemmli et al., 1970). After separation the gels were either stained with Coomassie blue (Sigma, St. Louis, USA) or treated for protein transfer to nitrocellulose membranes for antibody detection using polyclonal rabbit anti- β -casein antibodies affinity purified on highly pure native casein coupled to an affinity gel. The obtained results demonstrated efficient expression of recombinant human β -casein in yeast. The samples were co-electrophoresed with purified native human β -casein (obtained by acid precipitated human native β -casein further purified by ion exchange chromatography (Rowland, 1938, Kunz and Lönnerdal, 1990)) and the recombinant molecules migrated in an identical manner, indicating correct and intact protein (Figure 7).

The expression levels were estimated to be more than 50 mg per litre of culture of recombinant human β -casein.

EXAMPLE 4

20 Expression of recombinant human β -casein in transgenic mice

A plasmid containing the murine whey acidic protein, WAP, gene as a 7.2 kb genomic fragment cloned at EcoRI sites was obtained from Dr. Lothar Hennighausen (Campbell et al., 1984). This plasmid was digested with EcoRI and KpnI and separated on agarose gel to obtain the approximately 2.6 kb upstream regulatory element. The plasmid was also cleaved with SalI and EcoRI and separated on agarose gel, and the fragment containing the third exon, the last intron, the fourth exon and downstream positioned mRNA processing signals as an approximately 2.3 kb large DNA fragment. These two fragments were cloned into a plasmid constructed as a fusion of pUC19 and pUC9 to eliminate the restriction sites in the polylinker except for the EcoRI site, together with a human cDNA fragment, generating pS 88.

In order to facilitate cloning of human β -casein to the WAP regulatory elements by addition of KpnI and Sall sites and to introduce a sequence identical to the murine 5' untranslated upstream of the start codon, the original cDNA clone, pS 21, was modified by using polymerase chain reaction. Two PCR primers were synthesized, SYM 2044 and SYM 2045, with the following sequences:

SYM 2044 5'-CGGGTACCCTAAAGGACTTGACAGCCATGAAGGTCCTCATCCTCGCCTGCCTGG-3'

SYM 2045 5'-CGGTCGACTTAGACACTAATGGGGTTATGAACTG-3'

10 The plasmid pS 21 was used as template in a polymerase chain reaction with the two primers. The obtained 698 bp fragment encoding human β -casein with a KpnI site in the 5' end and a Sall site in the 3' end and a modified 5' untranslated sequence was then cloned into the approximately 7.5 kb KpnI and Sall digested pS 88 fragment and analyzed by DNA sequencing. The resulting expression vector was designated pS 133 (Figure 8). The sequence of the KpnI and Sall PCR generated fragment is listed in Figure 9.

The expression vector pS 133 is capable to mediate stage and tissue specific expression of recombinant human β -casein in transgenic animals. The regulatory elements direct gene expression to the mammary gland during lactation, thereby allowing isolation of the heterologous protein from milk.

Before injection of the expression vector into a pro-nucleus of a fertilized egg, pS 133 was digested with EcoRI and the WAP β -casein fragment was isolated on agarose gel followed by electroelution. The eluted DNA was precipitated and redissolved in 10 mM Tris (pH 7.5) and 0.1 mM EDTA for microinjection.

The experimental procedures employed to obtain transgenic animals is described in Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1986.

Transgenic mice were identified by analysis of DNA which has been prepared from excised tail samples. The tissue samples were incubated with proteinase K and phenol/chloroform extracted. The isolated DNA was used in polymerase chain reactions with primers which amplify specific fragments if the heterologous introduced DNA representing the expression vector fragment is present. The animals were also analyzed by DNA hybridization experiments to confirm PCR data and to test for possible rearrangements, structure of the integrated vector elements and to obtain information about the copy number of integrated vector elements.

In one set of experiments, 20 mice were analyzed with the two methods and the results demonstrated that 11 mice were carrying the heterologous DNA vector element derived from pS 133. The result from the PCR analysis and the hybridization experiments were identical. No rearrangements were observed from the hybridization results, and the number of integrated vector copies shows large variation, between 1 and 40 vector copies per cell.

The mice identified to carry vector DNA element, founder animals, were then mated and the F1 litter was analyzed for transgene by the same procedures.

Milk samples were collected from anesthetized founder animals treated with oxytocin to induce lactation and analyzed for the presence of recombinant human β -casein. This was done by SDS-PAGE, transfer to nitrocellulose membranes and incubation with polyclonal antibodies generated against native human β -casein. The obtained results demonstrated expression of recombinant human β -casein in milk from transgenic mice (Figure 10).

Stable lines of transgenic animals are generated.

In a similar manner, other transgenic animals such as cows or sheep capable of expressing human β -casein may be prepared.

SUBSTITUTE SHEET

EXAMPLE 5

Purification of recombinant β -casein from *E. coli*

E. coli cells expressing recombinant human β -casein and prepared as described in Example 2 above were separated from the culture by centrifugation and the resulting pellet was freeze-dried and thawed several times so as to disrupt the cells. Other methods for cell disruption includes osmotic shock, pressure change and sonication. After an additional centrifugation, the supernatant was collected and ammonium sulphate at varying concentration starting from 0.05 M was added so as to precipitate β -casein.

SDS-polyacrylamide gels (10-17%) (Figure 11A) as well as liquid chromatography (Figure 12) were run to analyze purity and Western blots (Figure 11B) to analyze identity. Antibodies used in Western blots were raised in rabbit using highly purified native human β -casein, purified from milk as described in Example 2 above, and purified by affinity chromatography on immobilized human β -casein. No cross-reactivity was seen with host proteins.

DEPOSITION

Plasmid DNA, designated pS 21, pS 26, pS 28, pS 133 and pS 232, has been deposited in the collection of Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-3300 Braunschweig, Germany, on August 19, 1991 in accordance with the provision of the Budapest Treaty and identified there by accession numbers DSM 6653, DSM 6654, DSM 6655, DSM 6656, and DSM 6657, respectively.

CLAIMS

1. A DNA sequence encoding a polypeptide comprising the amino acid sequence shown in Figure 1 or

an analogue of said DNA sequence which

- 5 1) hybridizes with the DNA sequence shown in Figure 1 under stringent hybridization conditions or
- 2) encodes a polypeptide, the amino acid sequence of which is at least 85% homologous with the amino acid sequence shown in Figure 1, or

- 10 3) constitutes an effective subsequence of said DNA sequence,

which analogue encodes a polypeptide having either the calcium binding activity of human β -casein, or opioid activity, or angiotensin converting enzyme (ACE) inhibitory activity, or a combination of any two or three of these activities.

- 15 2. A DNA sequence which comprises a human β -casein gene or an analogue of said DNA sequence which

 1) hybridizes with the DNA sequence shown in Figure 1 or a specific part thereof under stringent hybridization conditions,

- 2) encodes a polypeptide, the amino acid sequence of which is at least 85% homologous with the amino acid sequence shown in Figure 1, or
- 20 3) constitutes an effective subsequence of said DNA sequence,

which analogue encodes a polypeptide having either the calcium binding activity of human β -casein, or opioid activity, or angiotensin converting enzyme (ACE) inhibitory activity, or a combination of any two or three of these activities.

25

3. A DNA sequence according to claim 1 which comprises the DNA sequence shown in Figure 2.
4. A DNA sequence according to any of claims 1-3 encoding a polypeptide comprising amino acid sequence 1-210 shown in Figure 1.
5. A modified DNA sequence which differs from a DNA sequence as defined in any of claims 1-4 in that at least one nucleotide has been deleted, substituted or modified or at least one additional nucleotide has been inserted so as to result in a DNA sequence which encodes a polypeptide having a calcium-binding which is similar to, increased or decreased as compared to the calcium-binding activity of human β -casein, or a polypeptide having ACE-inhibitory activity, or a polypeptide having opioid activity.
6. Recombinant human β -casein comprising a polypeptide having the amino acid sequence shown in Figure 1 or a variant thereof having the calcium-binding activity of human β -casein or a subsequence of said amino acid sequence or variant having ACE-inhibitory activity or opioid activity.
7. A recombinant polypeptide encoded by a DNA sequence as claimed in any of claims 1-5.
8. A polypeptide according to claim 7 comprising the amino acid sequence shown in Figure 1.
9. A polypeptide according to claim 8 comprising amino acid sequence 1-210 shown in Figure 1.
10. A polypeptide according to any of claims 6-9 in which at least one amino acid residue has been substituted with a different amino acid residue and/or in which at least one amino acid residue has been deleted or added so as to result in a polypeptide comprising an amino acid sequence being different from the amino acid sequence shown in Figure 1 and having a similar, increased or decreased calcium-binding activity as compared to the calcium-binding activity of human β -

casein, or having ACE-inhibitory activity, or having opioid activity, or having a combination of any two or three of these activities.

11. A recombinant polypeptide according to any of claims 6-10 in which at least one amino acid residue has been modified by post-translational modification.

12. A polypeptide according to claim 11 in glycosylated and/or phosphorylated form.

13. A replicable expression vector which carries and is capable of mediating the expression of a DNA sequence encoding human β -casein.

14. A vector according to claim 13 in which the DNA sequence encoding human β -casein is as defined in any of claims 1-5.

15. A cell harbouring a vector as defined in any of claims 13 or 14.

16. A cell according to claim 15 which is a prokaryotic, a unicellular eukaryotic cell or a cell derived from a multicellular organism.

17. A method of producing a polypeptide according to any of claims 6-12, comprising

inserting a DNA sequence as defined in any of claims 1-5 in a vector which is able to replicate in a specific host cell, introducing the resulting recombinant vector into a host cell and growing the resulting cell in or on an appropriate culture medium under appropriate conditions for expression of the polypeptide and recovering the polypeptide.

18. A polypeptide according to any of claims 6-12 produced by a method as defined in claim 24.

19. A plasmid selected from the plasmids designated pS 21, pS 26, pS 28, pS 133 and pS 232 all of which have been deposited on 19 August, 1991 with the collection of Deutsche Sammlung v n Mikr organismen und Zellkulturen GmbH (DSMZ) under the accession numbers DSM 6653, DSM

6654, DSM 6655, DSM 6656, and DSM 6657, in accordance with the provisions of the Budapest Treaty.

20. A mammalian expression system comprising a DNA sequence encoding human β -casein inserted into a gene encoding a milk protein of a non-human mammal so as to form a hybrid gene which is expressible in the mammary gland of an adult female of a mammal harbouring said hybrid gene so that human β -casein is produced when the hybrid gene is expressed.
21. An expression system according to claim 20 in which the DNA sequence encoding human β -casein is a DNA sequence as defined in any of claims 1, 2 or 4-7.
22. An expression system according to claim 20 or 21 in which the gene encoding a milk protein is selected from a whey acidic protein (WAP) gene.
23. A hybrid gene comprising a DNA sequence encoding human β -casein inserted into a gene encoding a milk protein of a non-human mammal, the DNA sequence being inserted in the milk protein gene in such a manner that it is expressible in the mammary gland of an adult female of a mammal harbouring the hybrid gene.
24. A non-human mammalian cell harbouring an expression system as defined in any of claims 20-22.
25. A mammalian cell according to claim 24 which is an embryo cell.
26. A method of producing a transgenic non-human mammal capable of expressing human β -casein, comprising injecting an expression system as defined in any of claims 20-22 into a fertilized egg or a cell of an embryo of a mammal so as to incorporate the expression system into the germline of the mammal and developing the resulting injected fertilized egg or embryo into an adult female mammal.

27. A method of producing a transgenic non-human mammal capable of expressing human β -casein and substantially incapable of expressing β -casein from the mammal itself, comprising

- 5 1) destroying the mammalian β -casein expressing capability of the mammal so that substantially no mammalian β -casein is expressed and inserting an expression system as defined in any of claims 20-22 into the germline of the mammal in such a manner that human β -casein is expressed in the mammal and/or
 - 10 2) replacing the mammalian β -casein gene or part thereof with an expression system as defined in any of claims 20-22.
28. A transgenic non-human mammal harbouring a DNA sequence encoding human β -casein.
29. A transgenic non-human mammal as claimed in claim 28 in which the DNA sequence is a DNA sequence as defined in any of claims 1-5.
- 15 30. A transgenic non-human mammal as claimed in claim 28 or 29 in which the DNA sequence is present in the germline of the mammal.
31. A transgenic non-human mammal as claimed in claim 30 in which the DNA sequence is present in a milk protein gene of the mammal.
- 20 32. A transgenic non-human mammal prepared by the method of claim 26 or 27.
33. A mammal according to any of claims 28-32 which is selected from the group consisting of mice, rats, rabbits, sheep, pigs and cattle.
34. Progeny of a transgenic mammal as defined in any of claims 28-33.
- 25 35. A method of preparing human β -casein comprising collecting milk from a mammal as claimed in any of claims 28-34 and recovering the human β -casein from the milk.

36. Milk from a non-human mammal comprising recombinant human β -casein.

37. Milk from a non-human mammal comprising a polypeptide as defined in any of claims 6-12.

5 38. Milk obtained from a transgenic mammal as defined in any of claims 28-33.

39. The use of a polypeptide according to any of claims 6-12 or 18 as a nutrient supplement.

10 40. The use according to claim 39 in which the nutrient supplement is included in an infant formula.

41. The use of a polypeptide according to any of claims 6-12 or 18 in the preparation of an infant formula.

42. An infant formula comprising recombinant human β -casein.

15 43. An infant formula comprising a polypeptide as defined in any of claims 6-12 or 18.

44. An infant formula prepared from milk as defined in any of claims 36-38.

45. A method of isolating a casein protein from a mixture of components comprising

20 adding ammonium sulfate in a concentration equal to or above 0.05 M to the mixture of components so as to precipitate the casein protein, and

separating the precipitated casein protein from the mixture of components.

46. A method of isolating recombinant human β -casein from bacterial or yeast cells in which it is substantially intracellularly produced comprising

5 separating the cells harbouring the recombinant β -casein from the culture medium,

rupturing the separated cells so as to release their content of recombinant human β -casein,

optionally removing cell debris from the mixture of ruptured cells,

10 and precipitating the released recombinant human β -casein by addition of ammonium sulfate in a concentration equal to or above 0.05 M so as to obtain a precipitated β -casein, and

isolating the precipitated β -casein.

15 47. A method of isolating recombinant human β -casein from a culture of bacterial or yeast cells in which it has been substantially extracellularly produced comprising

removing the bacterial or yeast cells from the culture medium,

adding ammonium sulfate in a concentration equal to or above 0.05 M so as to obtain a precipitated β -casein, and

20 isolating the precipitated β -casein.

1/17

-15 -1 +1
M K V L I L A C L V A L A L A R E T I
cggatgaaggtcctcatcctcgctgcctgggtggctcttgctcttgcaagggagaccata
1 60

10 20
E S L S S S E E S I T E Y K K V E K V K
gaaagcctttcaagcagtgaggaatctattacagaatacaagaaagttgagaaggttaa
120

30 40
H E D Q Q Q G E D E H Q D K I Y P S F Q
catgaggaccagcagcaaggagaggatgaacaccaggataaaatctaccctctttccag
121 180

50 60
P Q P L I Y P F V E P I P Y G F L P Q N
ccacagcctctgatctatccattcggtgaacctatcccctatggttttcttccacaaac
181 240

70 80
I L P L A Q P A V V L P V P Q P E I M E
attctgcctcttgctcagcctgctgtgggtgctgcctgtccctcagcctgaaataatggaa
241-----> 300

<-----

90 100
V P K A K D T V Y T K G R V M P V L K S
gtccctaaagctaaagacactgtctacactaagggcagagtgatgcctgtccttaaact
301 360

110 120
P T I P F F D P Q I P K L T D L E N L H
ccaacgatacccttttttgaccctcaaattcccaaaactcactgatcttgaaaatctgcat
361

130 140
L P L P L L Q P L M Q Q V P Q P I P Q T
cttctctgcctctgctccagcccttgatgcagcaggtccctcagcctattcctcagact
480

Fig. 1A

2/17

150 160
L A L P P Q P L W S V P Q P K V L P I P
cttgcaacttccccctcagccccctgtggtctgttcctcagcccaaagtctgcctatcccc
481 -----
-----<-----
170 180
Q Q V V P Y P Q R A V P V Q A L L L N Q
cagcaagtgggtgccctaccctcagagagctgtgcctgttcaagcccttctgctcaaccaa
-----> 600
190 200
E L L L N P T H Q I Y P V T Q P L A P V
gaacttctacttaacccccaccacagatctaccctgtgactcagccacttgccccagtt
601 660
210
H N P I S V *
cataacccattagtggtctaagaagatttcaaagttaattttccctccttatttttgaat
661 720
tgactgagactggaaatatgatgccttttccgtctttgtatcacgttaccctcaaattaag
721 780
tatgtttgaatgagtttatatggaaaaaatgaactttgtccctttattttattttatatat
781 840
tatgtcattcatttaatttgaaatttgactcatgaactatttacattttccaaatcttaa
841 ----- 900
ttcaactagtagcacagaagttcaatactcatttggaaatgctacaaacatatcaaacat
901 960
atgtatacaaattgtttctggaattgtgcttatttttattttctttaagaatctatttcct
961 1020
ttccagtcatttcaataaattatttcttaagcataaaaaaaaaaaaaa
1021 1065

Fig. 1B

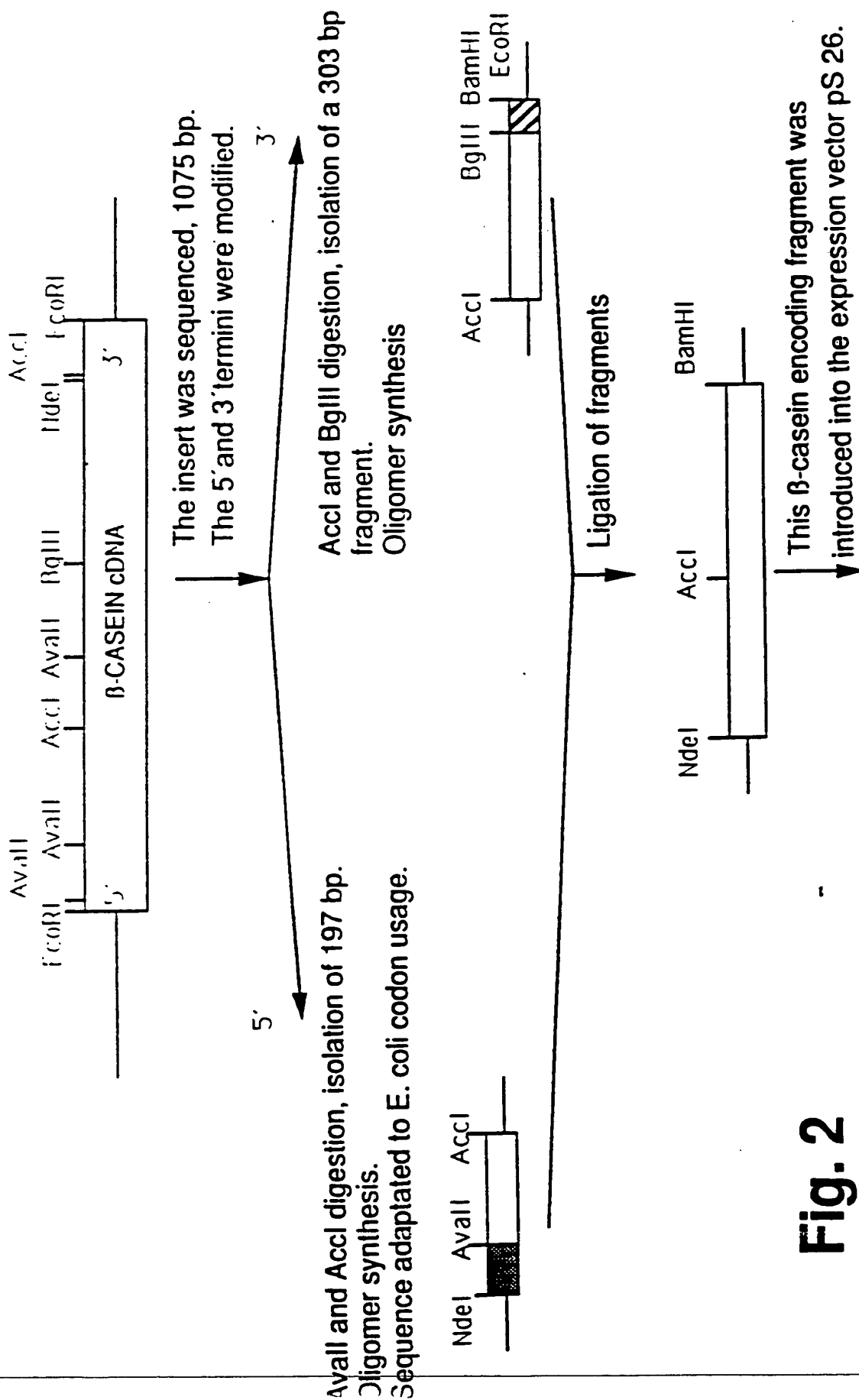


Fig. 2

4/17

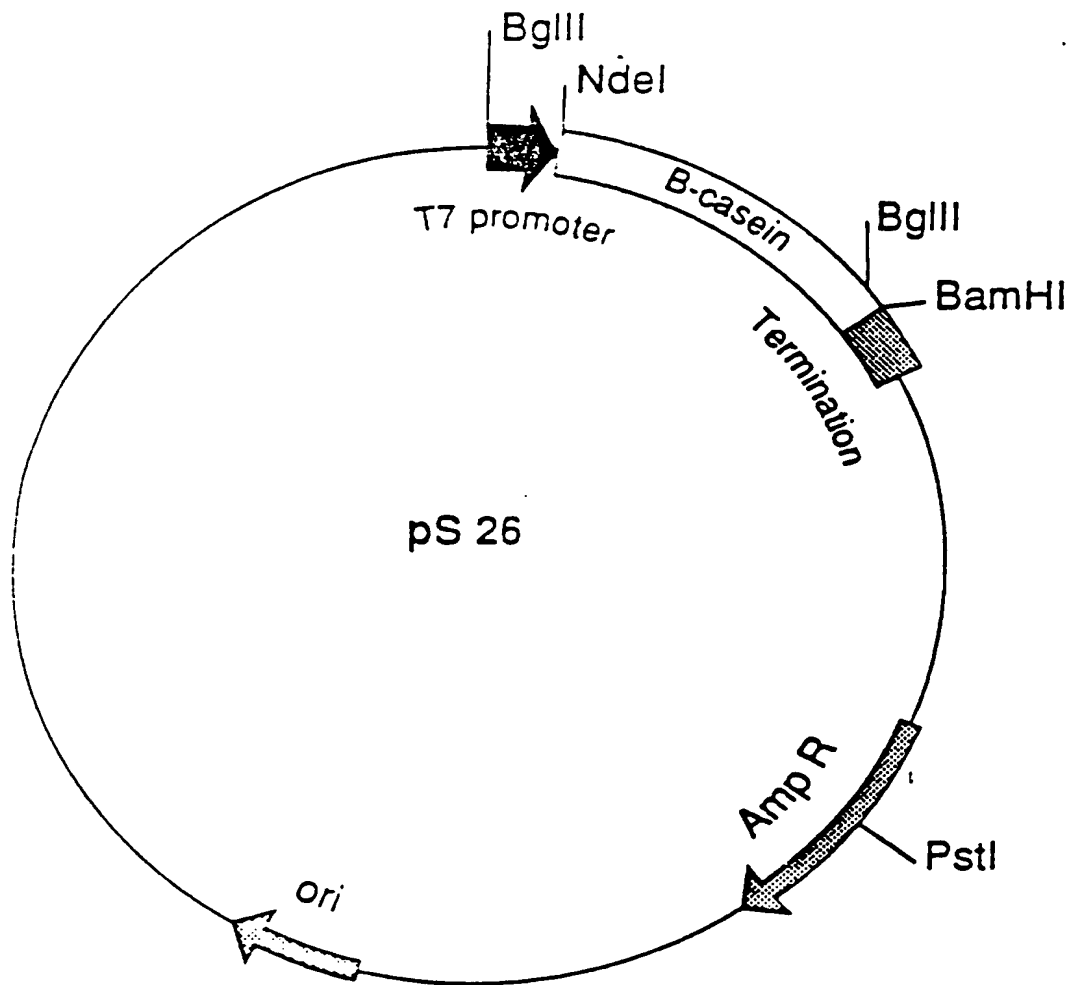


Fig. 3

5/17

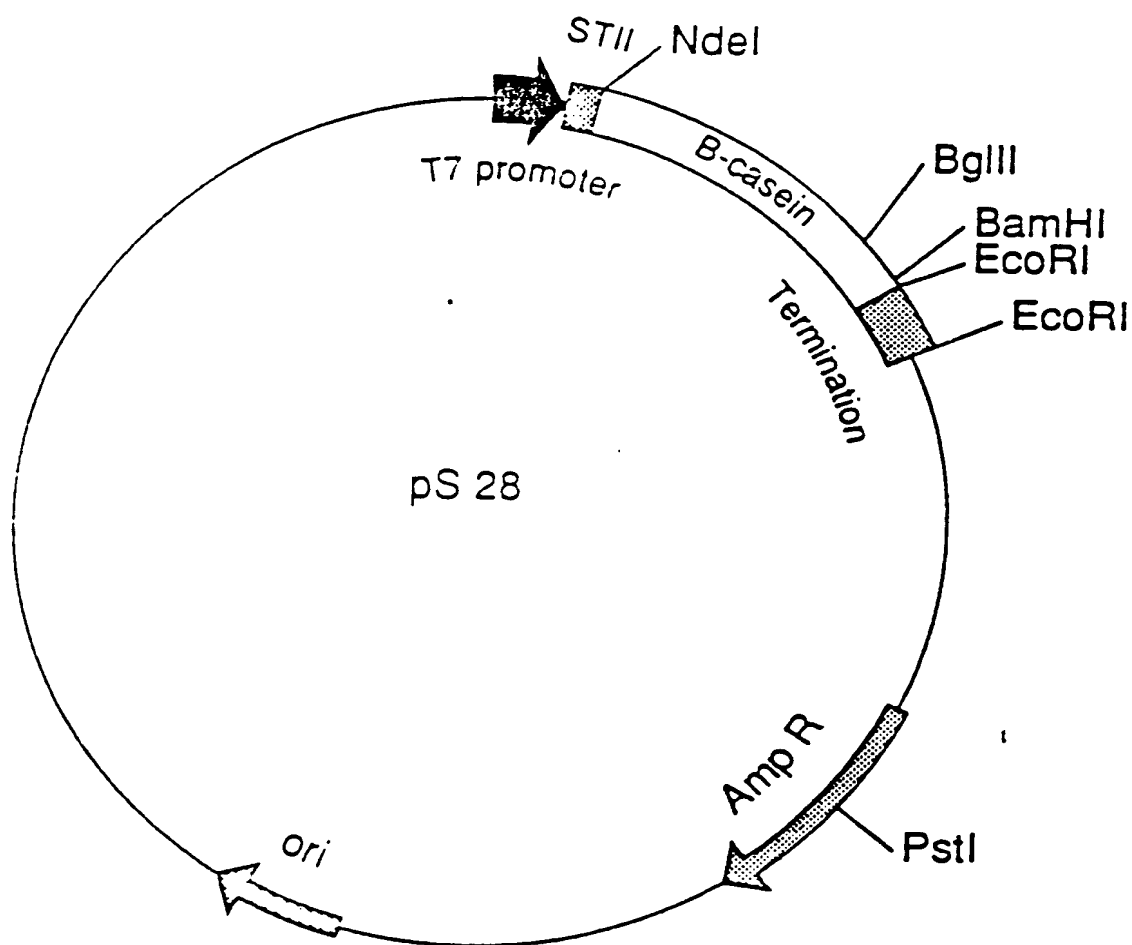


Fig. 4

6/17



Fig. 5A

7/17

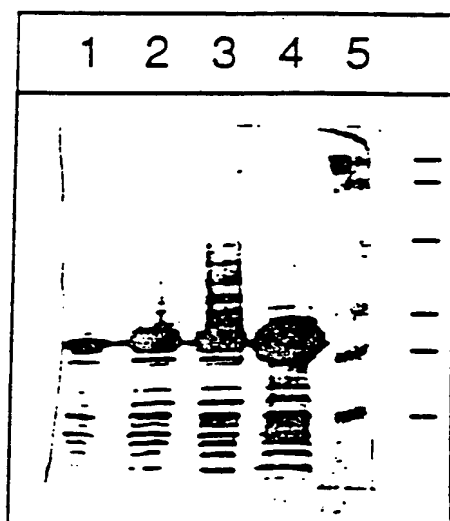


Fig. 5B

8/17

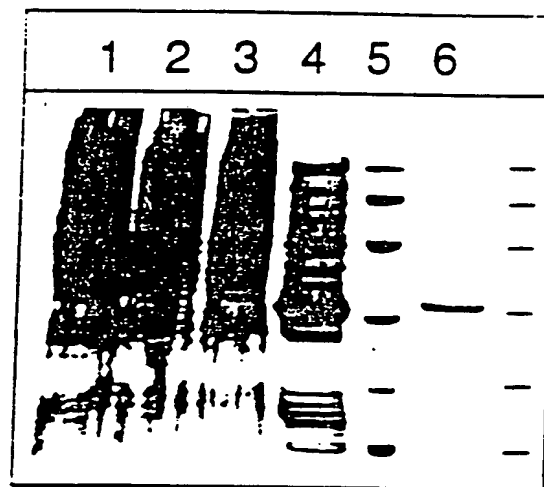


Fig. 5C

9/17

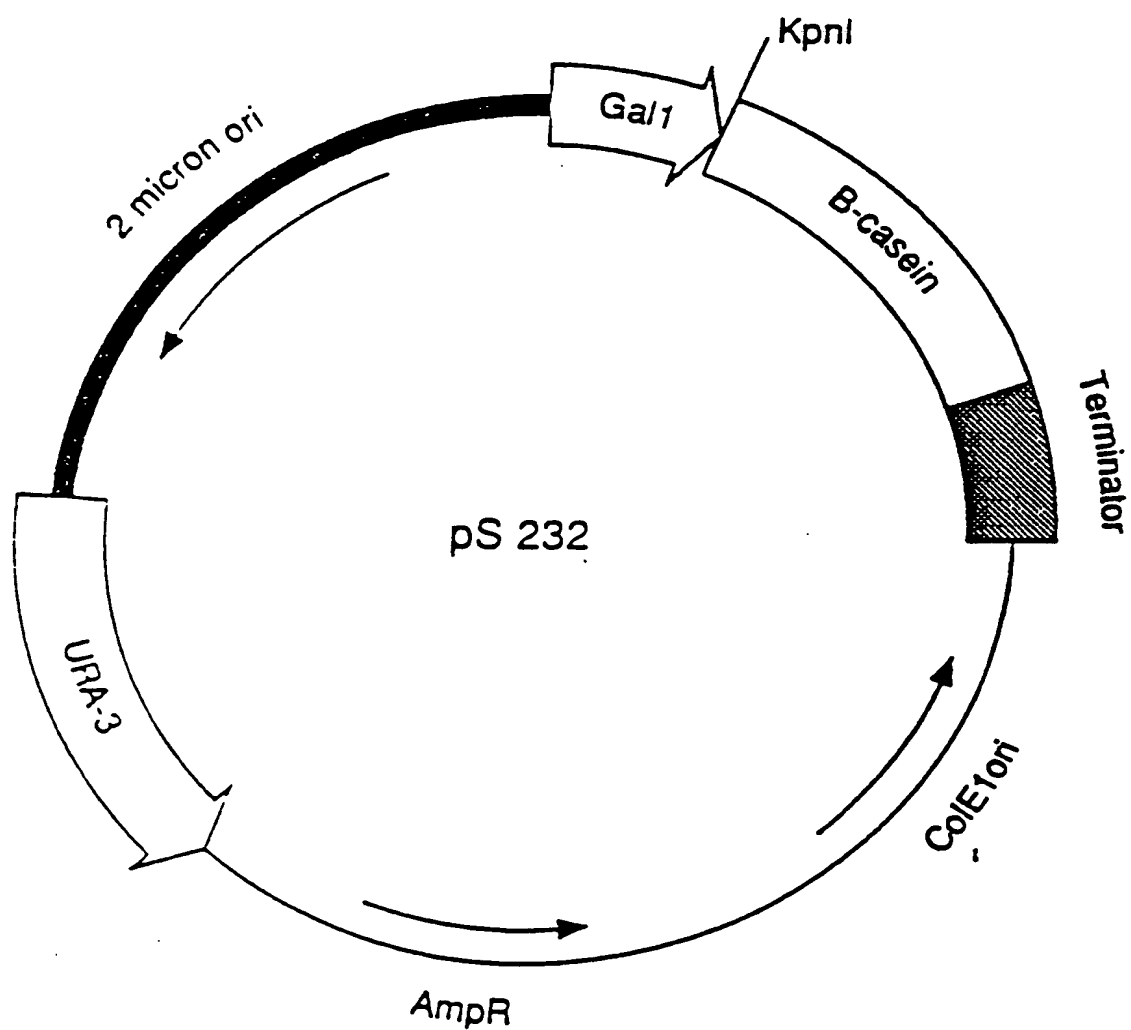


Fig. 6

10/17

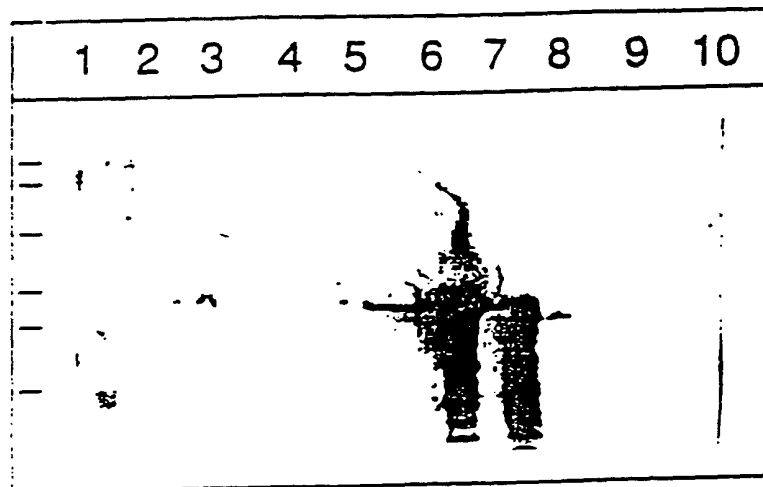


Fig. 7

11/17

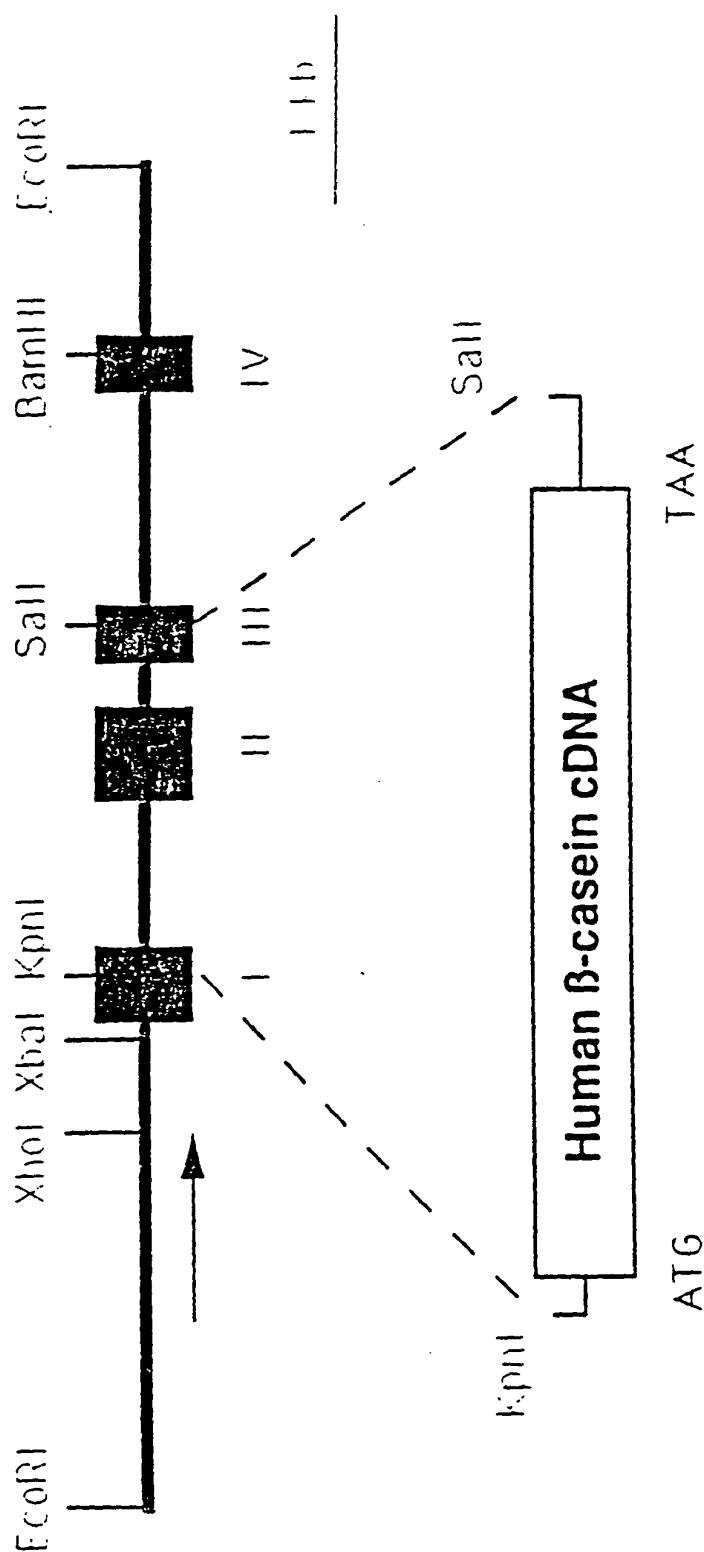


Fig. 8A

12/17

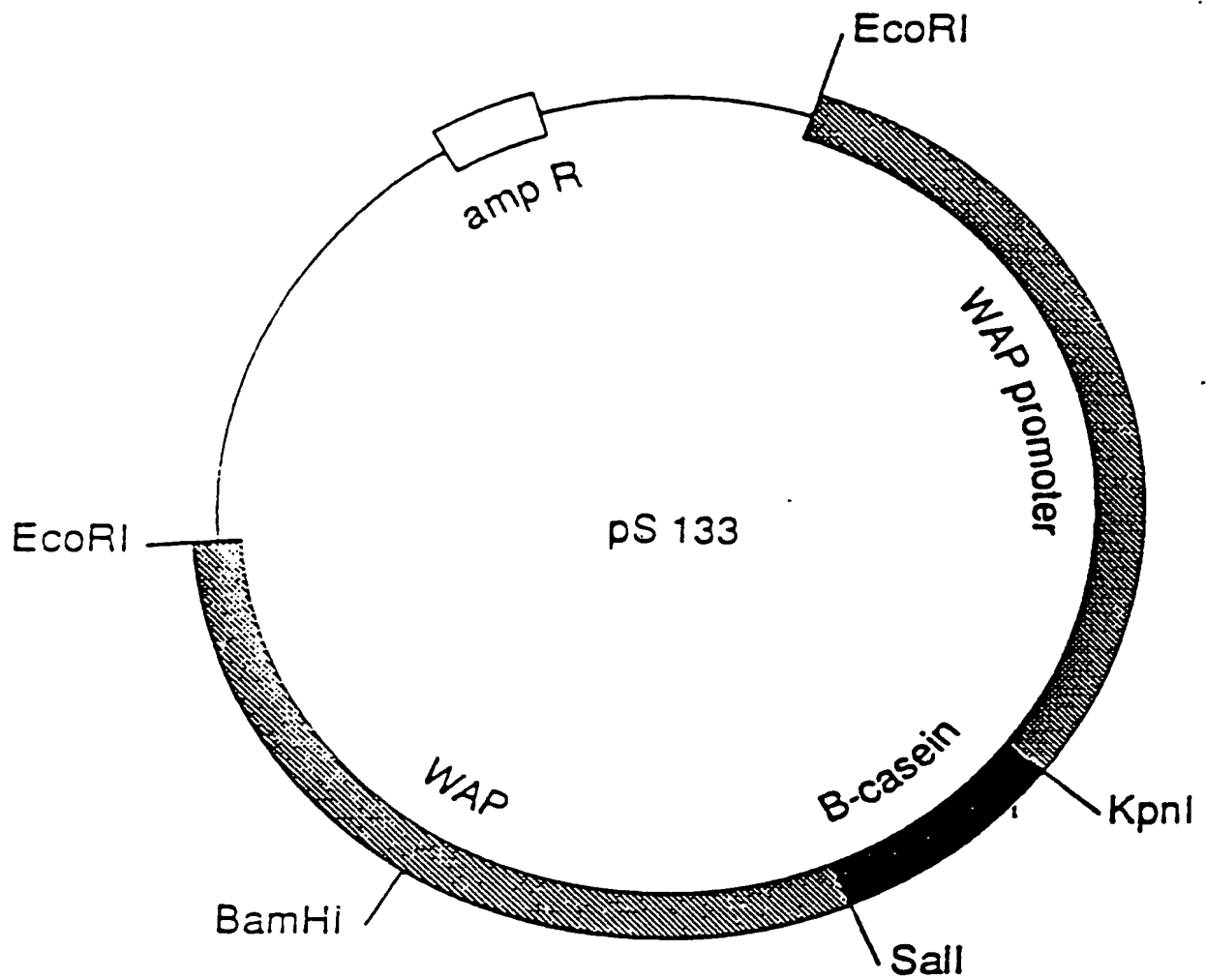


Fig. 8B

14/17

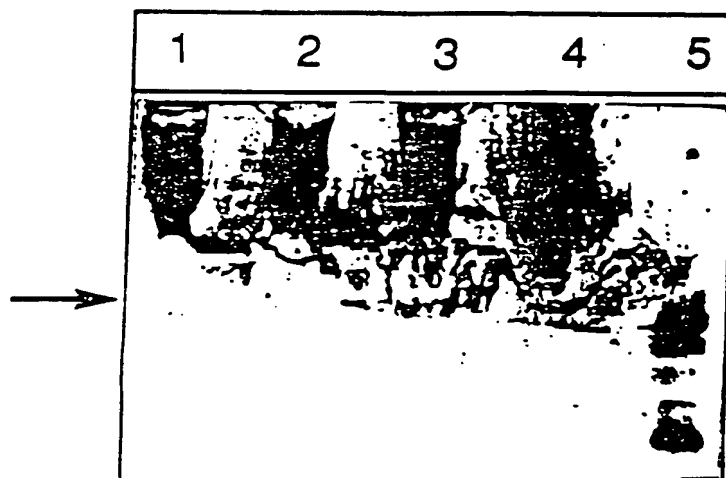


Fig. 10

15/17

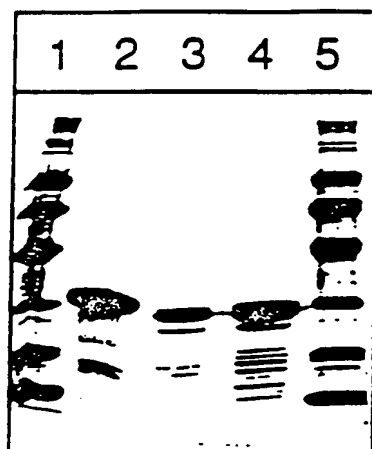


Fig. 11A

16/17

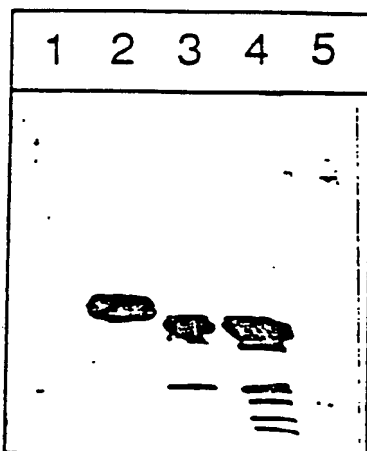


Fig. 11B

17/17

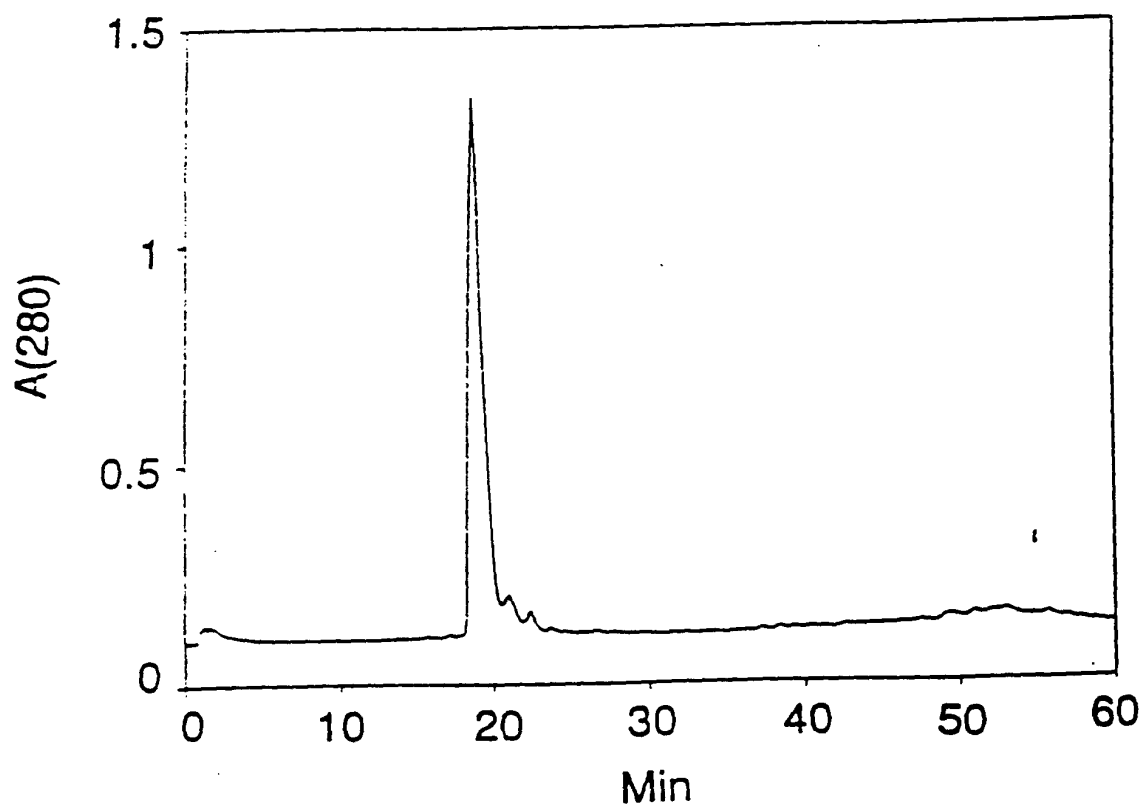


Fig. 12

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 46 line 19 of the description.

A. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet ☐

Name of depository institution *

Deutsche Sammlung von Mikroorganismen (DSM)

Address of depository institution (including postal code and country) *

Mascheroder Weg 1b
D-3300 Braunschweig
Federal Republic of Germany

Date of deposit *

August 19, 1991

Accession Number *

DSM 6657

B. ADDITIONAL INDICATIONS * (leave blank if not applicable). This information is continued on a separate attached sheet ☐

As regards the respective Patent Offices of the respective designated states, the applicant requests that a sample of the deposited microorganisms only be made available to an expert nominated by the requester until the date on which the patent is granted or the date on which the application has been refused or withdrawn or is deemed to be withdrawn

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE * (If the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS * (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later * (Specify the general nature of the indications e.g. "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)_____
(Authorized Officer)☐ The date of receipt (from the applicant) by the International Bureau is:

09 DECEMBER 1991

was

(09.12.91)

(Authorized Officer)

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/DK 91/00233

I. CLASSIFICATION OF SUBJECT MATTER (in several classification symbols apply, indicate all)⁹

According to International Patent Classification (IPC) or to both National Classification and IPC

Int. Cl. 5	C 12 N 15/12	C 12 P 21/02	C 07 K 13/00
C 12 N 1/21	C 12 N 1/19	C 12 N 15/03	C 12 N 15/85
A 23 C 9/14	A 23 C 11/00		

II. FIELDS SEARCHED

Minimum Documentation Searched⁷

Classification System

Classification Symbols

Int. Cl. 5	C 12 N	C 12 P	C 07 K
------------	--------	--------	--------

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched⁸

III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹

Category ¹	Citation of Document ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	FEBS Letters, volume 269, no. 1, August 1990, Elsevier Science Publishers B.V. (Amsterdam, NL) B. Lönnerdal et al.: "Cloning and sequencing of a cDNA encoding human milk beta-casein", pages 153-156, see the whole article especially figure 1 (cited in the application) ---	1, 2, 4, 19
Y	---	13-16, 20-38
X	WO/A, 9108675 (SLATTERY, Charles) 27 June 1991, see the whole document, especially examples I, II and the claims ---	13-17, 21, 46, 47
X	Faseb Journal, volume 4, no. 3, 1990, Bethesda MD A671, B. Lönnerdal et al.: "Cloning and sequencing of the gene for human beta-casein", abstract no. 2348, see abstract no. 2348 ---	1-2, 4-5
	-/-	

⁹ Special categories of cited documents: ¹⁰

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

26-03-1992

Date of Mailing of this International Search Report

27.04.92

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

[Handwritten Signature]

III. DOCUMENTS CONSIDERED TO BE RELEVANT

(CONTINUED FROM THE SECOND SHEET)

Category	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim
Y	---	20-38
Y	Proceedings of the National Academy of Sciences of the USA, volume 85, no. 16, August 1988, Biochemistry (Washington, DC, US) C.W. Pittius et al.: "A milk protein gene promoter directs the expression of human tissue plasminogen activator cDNA to the mammary gland in transgenic mice", pages 5874-5878, see the whole article	20-38
A	EP,A,0279582 (BAYLOR COLLEGE OF MEDICINE) 24 August 1988, see pages 13-19; claims	20-38
A	Journal of Dairy Science, volume 72, no. 16, 1989, (Champaign, Illinois, US) R.D. Bremel et al.: "Alteration of milk composition using molecular genetics", pages 2826-2833, see the whole article	20-38
Y	Methods in Enzymology, volume 185, 1990, Academic Press Ltd, (London, GB) F.W. Studier et al.: "Use of T7 RNA polymerase to direct expression of cloned genes", pages 60-89, see page 65, line 26 - page 69	13-16

ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO.

DK 9100233

SA 50921

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 14/07/92. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A- 9108675	27-06-91	AU-A- 7177391	18-07-91
EP-A- 0279582	24-08-88	AU-B- 618524	02-01-92
		AU-A- 1178488	18-08-88
		JP-A- 63309192	16-12-88

INTERNATIONAL SEARCH REPORT

PCT/DK91/00233

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
see PCT/ISA/206 mailed on 15.04.92

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims: it is covered by claims Nos.:

1-5, 13-17, 19-38, 46-47

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

1. The first part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

2. The second part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

3. The third part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

4. The fourth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

5. The fifth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

6. The sixth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

7. The seventh part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

8. The eighth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

9. The ninth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.

10. The tenth part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the office to which the person has been appointed.